



Guidelines

Document reader

Overview

The guidelines are specifically for users who are using the document reader. The manual provides detailed instructions regarding the use of the system and its user interfaces. The necessary background information and definitions are also provided for better comprehension. In addition, we offer further support via a wide range of training courses at HABEL or on your own premises.

Syntax conventions used in this manual

Representation	Meaning
<u>HABEL ANNOTATIONS</u>	Indication that the function described here constitutes an additional module that may possibly not be included in the scope of the system.
	A symbol is used to denote (warning) instructions and/or information to be observed.
	A symbol is used to denote background information and tips.

Discrepancies

The screenshots may differ in detail from your HABEL[®] document management program, as functions may be included that have not been purchased and/or activated for your system. Generally it is easy to add to your system. Please contact your consultant for details.

Note:

All rights reserved. No part of this manual may be reproduced or electronically processed, copied or disseminated in any way (by printing, photocopying, microfilming or other process) without the prior written authorization of HABEL. The contents of the manual are subject to revision without notice on the part of HABEL.

Liability and warranty

This manual has been compiled and checked with the greatest possible care. However, the possibility of errors cannot be completely excluded. HABEL does not accept any legal responsibility, liability or warranty for incorrect information and the consequences thereof.

HABEL GmbH & Co. KG
Untere Hauptstraße 1
D-78604 Rietheim-Weilheim
Telephone: +49 (0) 74 61 / 93 53 – 0
Fax: +49 (0) 74 61 / 93 53 – 99
www.habel.de

Copyright © 2008 – 2011 by HABEL / As of 02.01.2012 / Version 2

HABEL Dokumentenmanagement GmbH
Swiss branch
Rheinstrasse 36
CH-8212 Neuhausen am Rheinfall
Telephone: +41 (0) 52 / 6 74 81 – 51
Fax: +41 (0) 52 / 6 74 81 – 50
www.habel.ch

Table of contents

Versions of the document reader.....	1
Uses of the document reader	2
Document entry/recognition.....	7
1. General	7
2. Scanning documents	7
3. Recognition of document information	8
4. Form dossiers	9
5. Correct recognized data	12
6. The recognition of the creditor/debtor	13
7. Other commands in the program "Forming dossiers"	18
Create templates	22
8. Scan the document (first step).....	24
9. Save the document (second step).....	24
10. Create a name for the template (third step)	25
11. Assign the document as the first page of the template (fourth step)	27
12. Identify the supplier (fifth step)	28
13. Define the fields/columns to be read out (sixth step)	32
14. Save additional pages (seventh step)	42
15. Special case: Multiple anchor field	42
16. Special case: Multiple anchor fixed field	44
17. Special case: Dummy	46
18. Special case: Negation of values	46
19. Import/export folder	46

Versions of the document reader

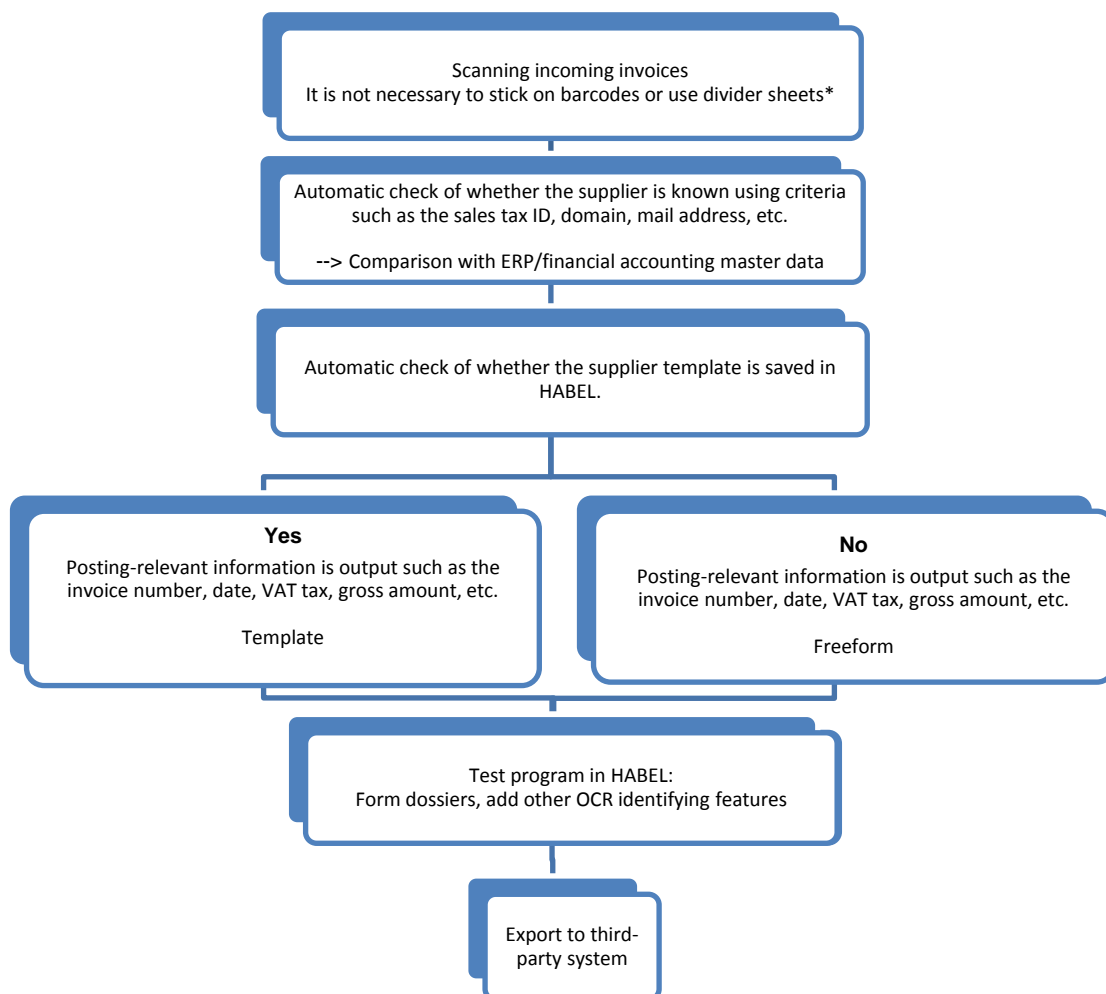
There are three versions of the document reader that have different functions. The right version for you depends on your requirements. The different uses of the document reader will be described below (of the third version in our case). The differences between the versions are indicated in the following table:

Function	DOCUMENT READER BL50	DOCUMENT READER BL100	DOCUMENT READER
OCR	yes	yes	yes
ICR	yes	yes	yes
Max. volume of templates	Maximum of 50	Maximum of 100	Unlimited
Read out the data with the templates	yes	yes	yes
Affix barcodes to separate the dossiers.	yes	yes	No, but possible
Read out	Maximum data for 10 headers and footers	Maximum data for 20 headers and footers	Unlimited
Position recognition	no	no	yes
Automatic invoice verification	no	no	Yes with additional module
Scripts are used	no	no	yes
- Client recognition	no	no	yes
- Plausibility checks	no	no	yes
- Document classification	yes	yes	yes
- Check the recognized data against external sources and other interpretations of the data	no	no	yes

Uses of the document reader

1. Read out document information and provide values

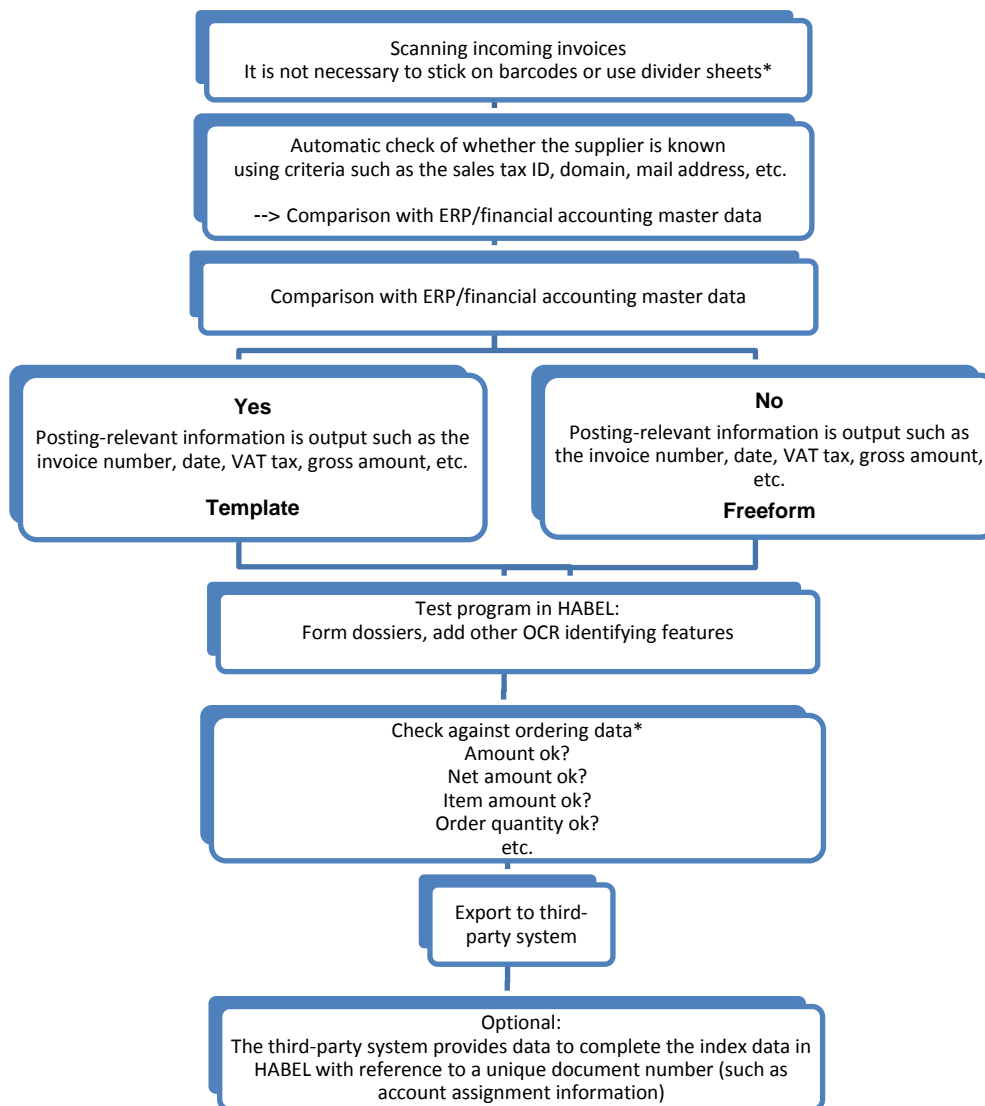
The document reader can be used to read out document information and provide values for further processing. The following is a graphic representation of an example of an incoming invoice:



**Not in connection with document reader BL50 / document reader BL 100*

2. Read out document information, interpretation, check the data and provide values

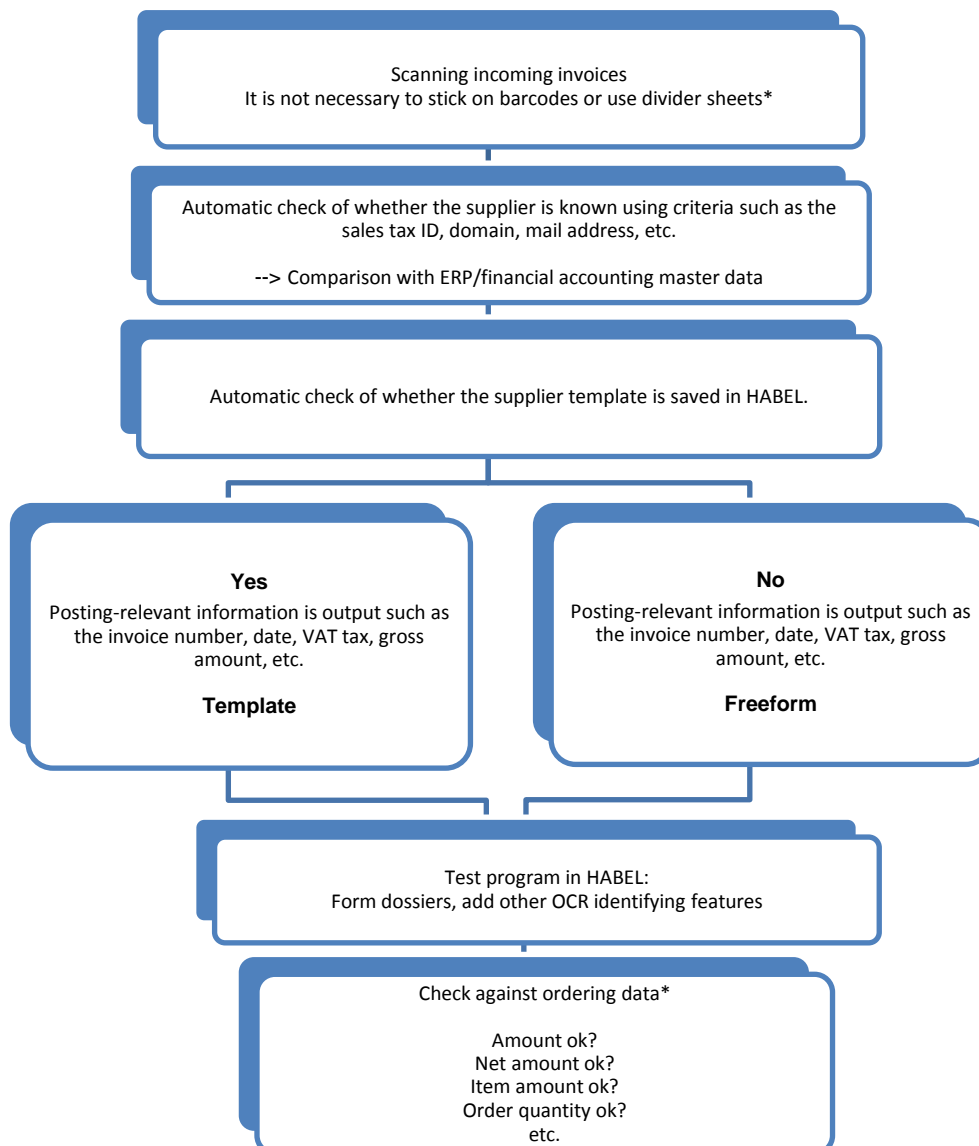
In addition, data such as currency abbreviations can be interpreted and added. To automate processes such as releases for posting, the data can be checked against ERP and financial accounting data. The following is a graphic representation of an example of an incoming invoice:



*Not in connection with document reader BL50 / document reader BL 100

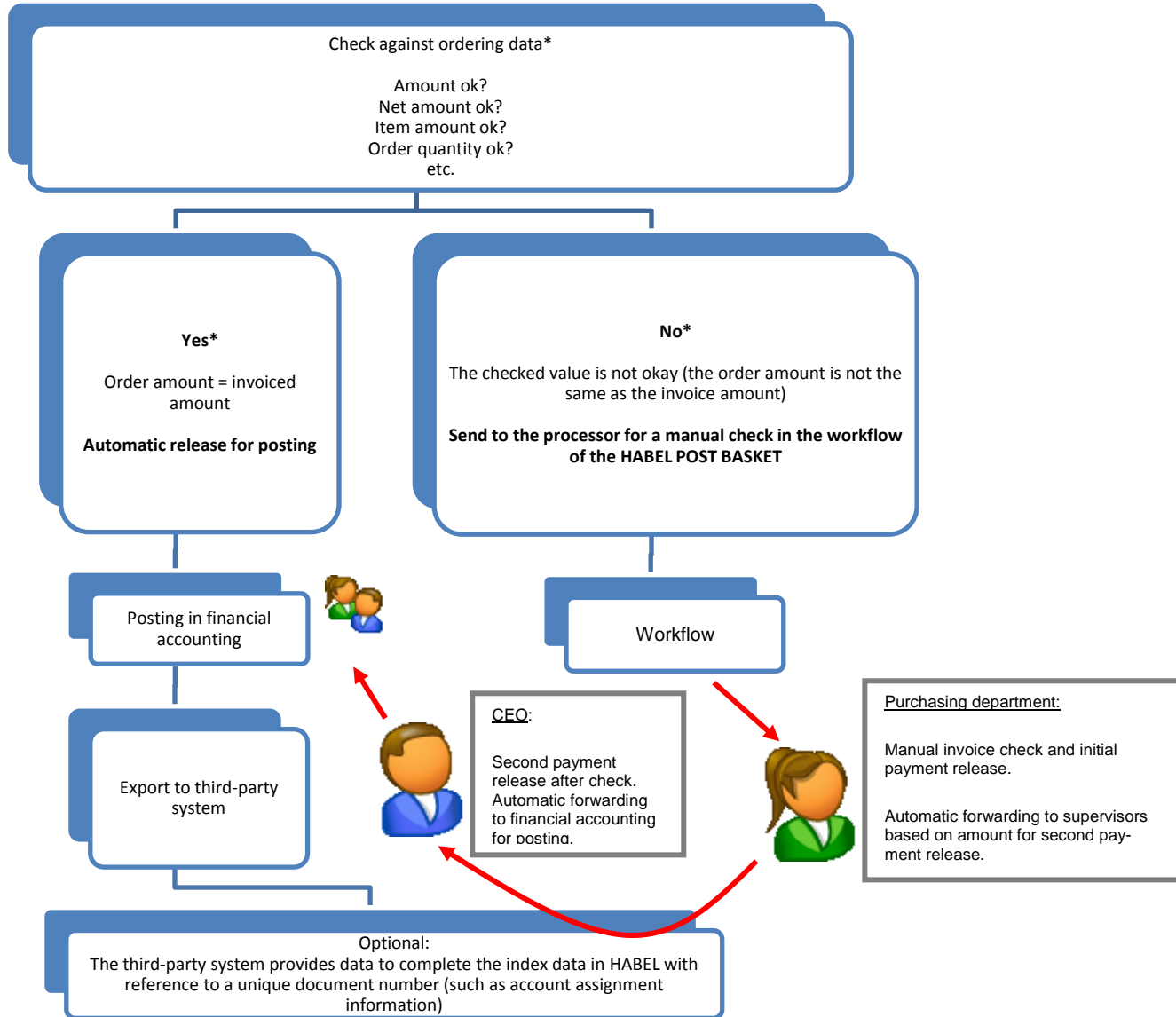
3. Read out document information, interpretation, check data, read out data in the post basket and provide values

If incongruities are identified when checking the data, the data can be manually processed in the post basket. Likewise, the post basket can be used for basic editing (reply). The following is a graphic representation of an example of an incoming invoice:



*Not in connection with document reader BL50 / document reader BL 100

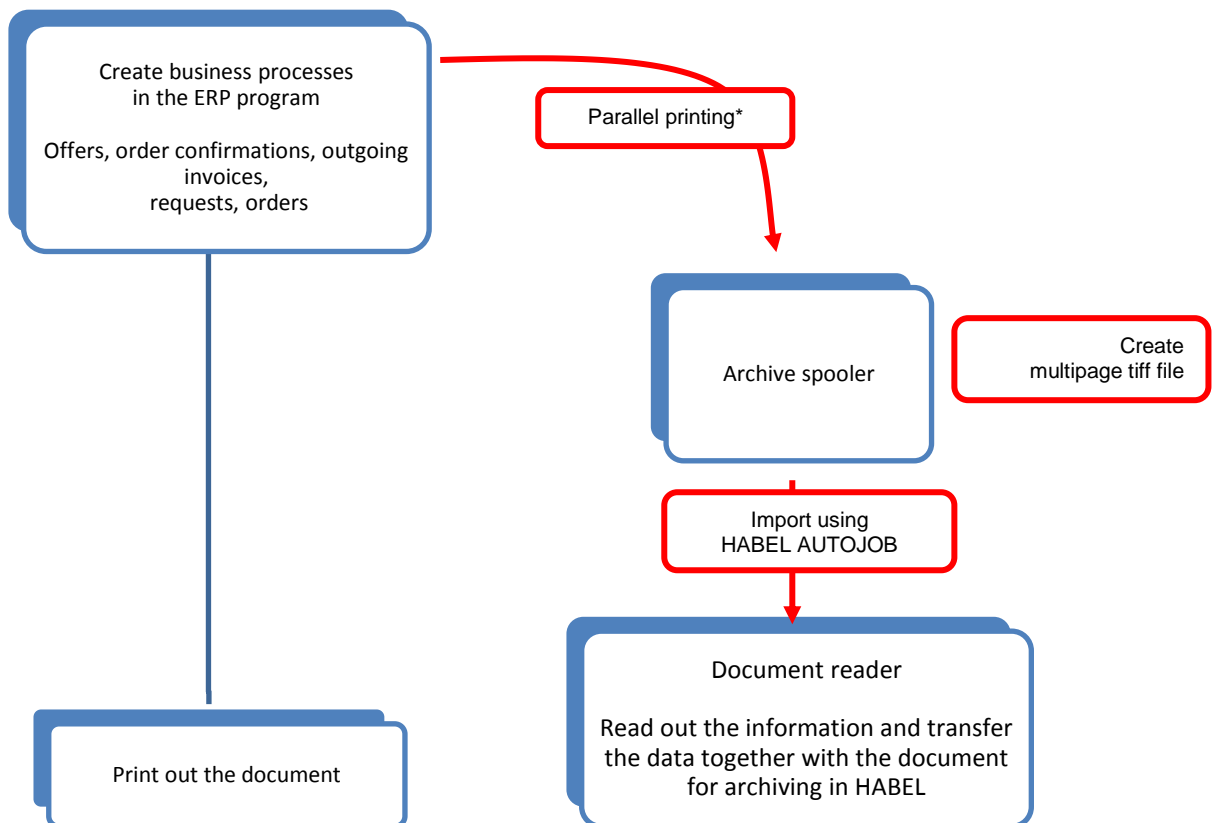
Continued



*Not in connection with document reader BL50 / document reader BL 100

4. Automatic archiving of outgoing documents

The document reader can also be used to archive outgoing documents that are provided without index data. The following graphic shows the process:



**Advantage: In this printing process, the digital level is never left. The quality of the document remains constant while it is being read out by the document reader, and the recognition rate is nearly 100%.*

Document entry/recognition

1. General

The document reader is used to read information directly from the documents when scanning them and to employ the information for processing. There are many different uses for the document reader; for example, it can scan and edit customer orders or incoming delivery notes. This module is frequently used to edit incoming invoices. The descriptions and hence associated terms in this manual refer to the editing of incoming invoices; however, they are equally applicable to other areas.

Various individual programs can be used in the document reader. Users primarily employ KOFAX (hpkfx100), the familiar scanning program, and the program for forming dossiers and training the document reader (hphab401). Other programs are available to the administrator, which will be described in a different juncture. The programs can be found in the directory of the HABEL server under \GEDOSOD\32-BIT\Program name.

In addition, an OCR license (ABBYY) is necessary to read out the raw text of the documents. A license must be available wherever the document reader is used.

2. Scanning documents

The documents are scanned in a stack using the KOFAX scanning program. It is not necessary to separate the transactions by barcodes or divider sheets. The documents that belong to a transaction are defined using a criterion (generally the document number). The document type is selected that was set up for processing using the document reader. The fields are also specified that need to be entered for scanning.

Company:	01	HABEL GmbH & Co. KG
Document type:	0099	BELEGLESER DOKUMENTE (OCR)
Document date:	_/_/___	
Contact person (external):		

Individual Stack Next Quit

Figure 1: Scanning program

3. Recognition of document information

After scanning (the data has been transmitted), the raw text can be recognized. This is done at specific times or is alternately triggered after scanning (end, transmit). Alternately, the processor responsible for editing can be informed when recognition is finished.

The program (hphab400) can also be started manually.

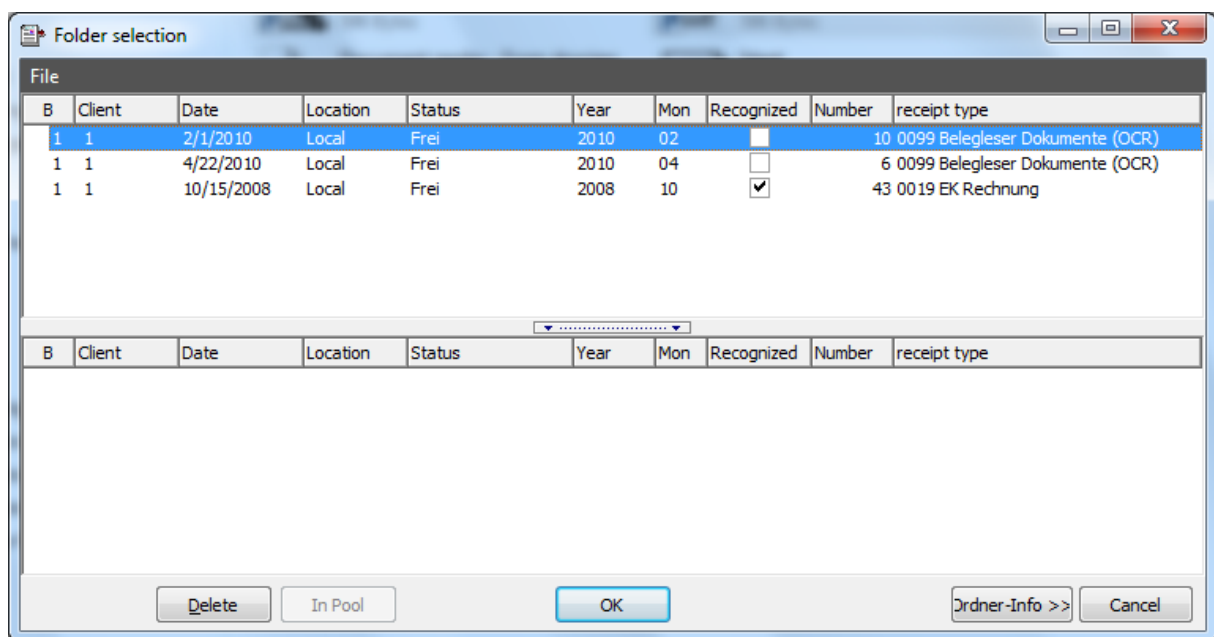


Figure 2: Folder selection

The scanned stack is shown in the top area with the compilation date, status, year, message if recognized, number, document type, etc. If the wrong stacks are scanned, they can be deleted by pressing the Delete button. These cannot be found in the archive. After the stack is selected, press the OK button and then *Recognize all* to start recognition.

The recognition of the document information involves the following steps:

- 1) Recognition of the creditor by means of freeform recognition
- 2) The document information is read out in freeform
- 3) Check if there is a template for the creditor, and read out the information using the template (if necessary, overwrite the values recognized in freeform recognition)

4. Form dossiers

The documents are displayed in the "Form dossiers" program (hphab401) for further editing. When selecting folders, you decide which stack to edit. The selected folders being processed or the stacks to be recognized are in the bottom area of folder selection.

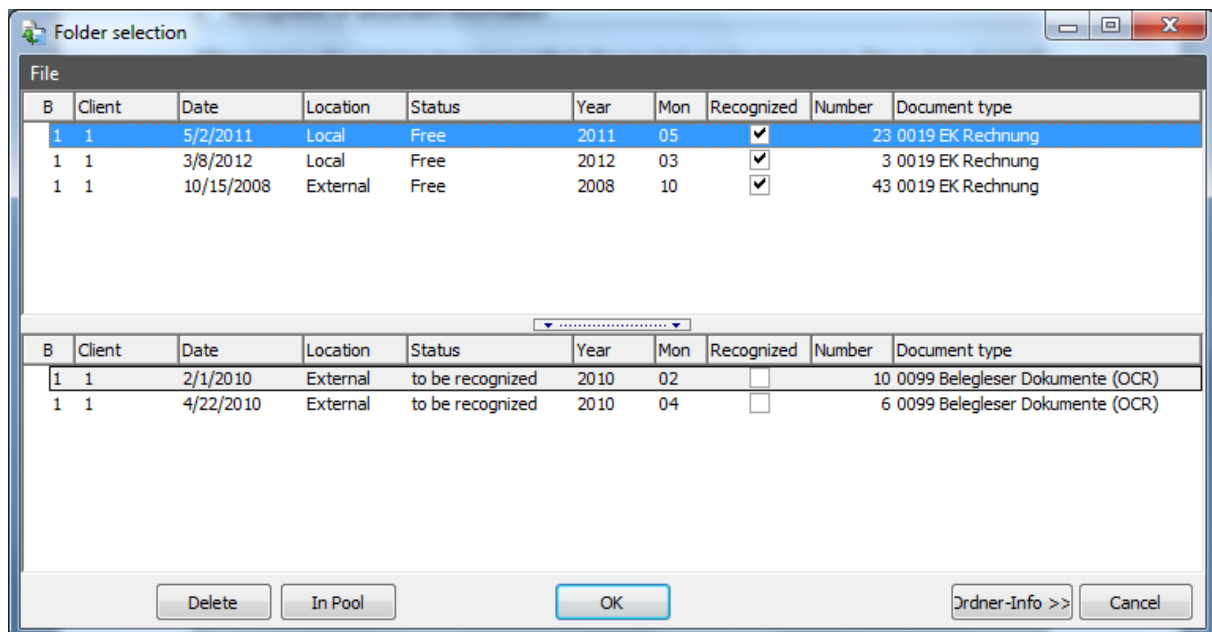
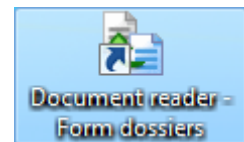


Figure 3: Folder selection for creating dossiers

Press the *Delete* button to delete the entire stack if it was mistakenly scanned. Remember: If a stack is deleted, the documents cannot be found in the archive even though data were transmitted after scanning. The documents are in the archive only after editing is finished with the document reader.

Note: Even if the recognition of the raw text has already occurred, the workplace at which the dossiers were formed requires access to OCR licenses (local ABBYY installation with access to the server). However, the page volume is not used at this point.

Assign receipts

File Viewer Action Tools - Program info -

1

Cosmetics International

HOTEL
Akademie Habel
Untere Hauptstraße 1
68004 Rheinheim-Weinheim

Rechnung

Belegnummer: 10887159 Kontakt: Max Mustermann
Belegdatum: 21.04.2011 Telefon: 07863698-4711
Kunde: 175339 Fax:
Ihr Zeichen: tel. Frau Nathalie Nasarek Email: max.mustermann@ada-cosmetics.com
Ihre USI-Hd-Nr.: Unsere USI-Hd-Nr.: DE252459565
Seite: 1 Unsere Steuer-Nr.: 140026859

Herzlichen Dank für Ihren Auftrag. Wir erlauben uns wie folgt zu berechnen:

Pos	Artikel	Menge	Einzelpreis EUR	Gesamtpreis EUR
10	FIBL030-100200-8 Body Lotion im Flacon Fidji 30ml HYDRO Basics	168 Stück	1,0500	176,40
20	PWBS300-100200-7 Hair + Body Shampoo 300ml in Dosierflacon press + wash HYDRO Basics	120 Stück	6,2500	750,00
30	PWLS300-100200-7 Liquid Cream Soap 300ml in Dosierflacon press + wash HYDRO Basics	60 Stück	6,2500	375,00
40	FIDG030-100200-7 Bath & Shower Gel im Flacon Fidji 30ml HYDRO Basics	336 Stück	0,9500	319,20
50	HWWB000-100320-1 Waschebeutel PE-LD mit Kordel 420 x 580 x 0,03 mm Accessories white	1.000 Stück	0,1500	150,00
60	HKDH009-100200-7 Duschhaube in trapezförmiger Körnung HYDRO Basics	200 Stück	0,1000	20,00
70	HKSS004-100200-7 Schuhschwamm in Dreieckskartongabe HYDRO Basics	800 Stück	0,1700	136,00
Übertrag				1.926,60

ADA Cosmetics International GmbH
Tel: +49 7803 998 400
Nassauer Straße 2 A Fax: +49 7803 998 410
71634 Keil, Germany info@ada-cosmetics.com
Hastach 1764 - 1767 Keil www.ada-cosmetics.com

Geschäftsinhaber: Werner B. Försching
Arbeitsgemeinschaft Amberg-Weiden, HRB 100 719
Deutsche Bank Offenburg BLZ 664 100 300 086 000 100
Commerzbank Baden-Baden BLZ 251 902 000 011 217 300
Forts Bank Kärnten BLZ 370 108 028 100 140 14 01

2

FA	Bezeichnung	Kd.-/Lief.-Name	External Document No.	Gesamt Netto	Gesamt Brutto	Bele
01	EK Rechnung	KOFAX Deutschland AG	2933743	297,56	354,10	
01	EK Rechnung	ADA Cosmetics GmbH	10887159			
01	EK Rechnung	ADA Cosmetics GmbH	10887159	2142,60	2549,69	
01	EK Rechnung	Alde Gott eG	135944	181,43	215,90	
01	EK Rechnung	Peter Lazic GmbH	-1	37,50	44,63	
01	EK Lieferschein	Alde Gott eG	136243			
01	EK Rechnung	Alde Gott eG	114770	272,77	324,60	
01	EK Gutschrift	ADA Cosmetics GmbH	10887160			
01	EK Gutschrift	ADA Cosmetics GmbH	10887160	2142,60	2549,69	
01	EK Rechnung	Abbyy Europe GmbH	2801488			
01	EK Rechnung	Abbyy Europe GmbH	2801488	264,00	314,16	
01	EK Rechnung	Richter Aluminium GmbH	47480			
01	EK Rechnung	Richter Aluminium GmbH	47480			
01	EK Rechnung	Richter Aluminium GmbH	47480			
01	EK Rechnung	Richter Aluminium GmbH	47480	35690,71	42471,94	
01	EK Lieferschein	Alde Gott eG	115640			

3

Kd.-/Lief. Nr: 42269
Belegart: 0019 Supplier Invoice
DTA-Code:
Fremdbeleg Nr: 10887159

Save Delete Cancel

4

Position	Fremdartikel Nr	Fremdartikelbezeichnung	Einzelpreis	Menge	Gesamtpreis
10	FIBL030-100200-8	Body Lotion im Flacon Fidji 30ml H1	1,0500	168	176,40
20	PWBS300-100200-7	Hair + Body Shampoo 300ml in Do	6,2500	120	750,00
30	PWLS300-100200-7	Liquid Cream Soap 300ml in Dosie	6,2500	60	375,00
40	FIDG030-100200-7	Bath & Shower Gel im Flacon Fidj	0,9500	336	319,20
50	HWWB000-100320-1	Waschebeutel PE-LD mit Kordel 42	0,1500	1000	150,00
60	HKDH009-100200-7	Duschhaube in trapezförmiger Kar	0,1000	200	20,00

5

Position: 10
Fremdartikel Nr: FIBL030-100200-8
Fremdartikelbezeichnung: Body Lotion im Flacon Fidji 30ml HYDRO Basics

Save Delete Neu Cancel

Recognize all Recognize **7** Export **8** Create processes **9** Combine Close

Copyright © by HABEL GmbH & Co. KG X:316 Y:3668

Figure 4: Form dossiers

After the folder is selected, the documents and recognized data are displayed in the program. The documents are displayed on the left **(1)**, the top right shows a hit list of the recognized dossiers **(2)**. In the area at the middle right, all the data of an entry are displayed. You can make changes which includes entering values, accessing the master data or deleting the entire record **(3)**. The same is true for positions. The positions are displayed in a table **(4)** and can be selected for individual processing (change, add and delete) **(5)**. The editing fields for the entire entry and the positions can be hidden or shown **(6)**.

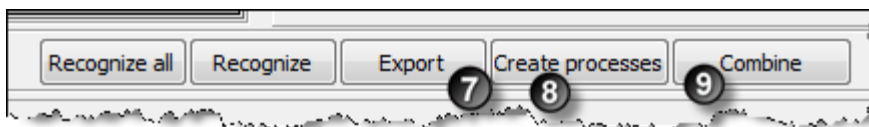


Figure 5: Form dossiers Footer

If the lines in the hit list have the same color and shading, the document reader has recognized them as a dossier. Each new dossier is shaded differently for the sake of clarity. Green dossiers are recognized by the document reader with reference to a specific criterion (such as the document number). If the entries are red, it cannot be determined whether the displayed page is the first page of the new dossier or if it belongs to the previous or following dossier.

If the entry is a single page and it is red, the criterion needs to be assigned to it (in this case, the document number). Select the entry, and press the *Join* (9) button. The criterion is requested.

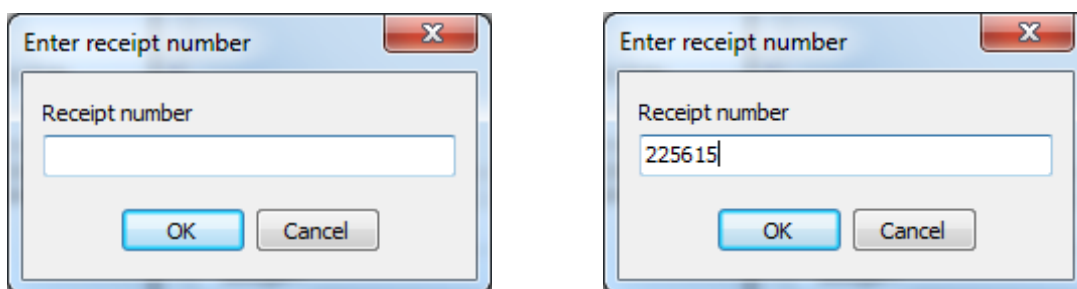


Figure 6: Enter the document number

After the document number is entered, press the *Form dossiers* button (8), and the entry appears as correct (marked green).

For multipage entries that are not recognized as a dossier, select the ones that belong together, and press the *Join* button (9). The correct document number is requested which can then be entered. Then press the *Form dossiers* button (8) to display the event as marked green.

If several document numbers were recognized for the entries that form a dossier, a selection window appears after the *Join* button (9) is pressed so that you can assign the correct document number.

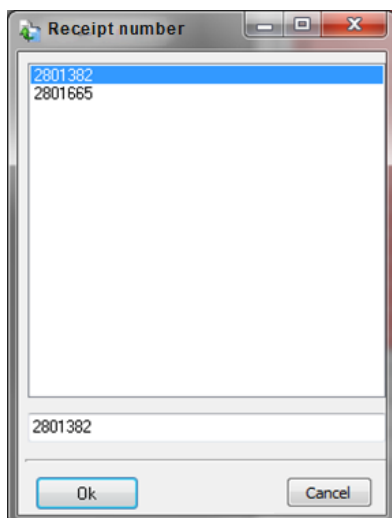


Figure 7: Assign document numbers

The document number that is selected is considered the correct document number, and it is assigned to all the entries within the dossier after the OK button is pressed. The dossier is displayed green after pressing the *Form dossiers* button (8).

Note: Even after joining the dossiers and pressing the *Form dossiers* button (8), the entry can still be marked red. This will happen if there are other entries before or afterward that cannot be correctly assigned. Of course, you can press the *Form dossiers* button (8) after all the entries have been joined.

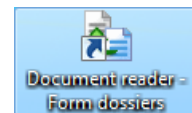
Once the dossiers have been formed, the recognized data can be exported for further processing (*Export* button (7)). Red entries are not exported.

One example of processing could be displaying the dossiers in the post basket that are incorrect according to saved rules and need to

be manually checked; for example, certain fields need to be filled in that were not filled in while recognizing the document. Likewise, you can additionally process the transfer of the recognized data to accounting for posting the dossiers. Or alternately, the data are fed into the workflow of an external program.

5. Correct recognized data

In the "Form dossiers" program (hphab401), recognized data can be manually corrected, or missing data can be added.



The screenshot shows the 'Assign receipts' window. On the left is a scanned receipt from Richter Aluminium. On the right is a data entry table with columns: FA, Bezeichnung, Kd.-Lief.-Name, External Document No., Gesamt Netto, Gesamt Brutto, and Sele. The table contains multiple rows of data, some of which are highlighted in red to indicate errors. A dialog box is open over the table, allowing for manual entry and correction of values. The dialog has fields for 'DTA-Code', 'Fremdbeleg Nr.', 'Document date', and 'Gesamt Brutto'. Annotations 1 through 5 are placed on the interface to guide the user through the correction process.

FA	Bezeichnung	Kd.-Lief.-Name	External Document No.	Gesamt Netto	Gesamt Brutto	Sele
01	EK Rechnung	KOFAX Deutschland AG	2933743	297.56	354.10	
01	EK Rechnung	ADA Cosmetics GmbH	10887159			
01	EK Rechnung	ADA Cosmetics GmbH	10887159	2142.60	2549.69	
01	EK Rechnung	Alde Gott eG	135944	181.43	215.90	
01	EK Rechnung	Peter Lacz GmbH	-1	37.50	44.63	
01	EK Lieferschein	Alde Gott eG	136243			
01	EK Rechnung	Alde Gott eG	114770	272.77	324.60	
01	EK Gutschrift	ADA Cosmetics GmbH	10887160			
01	EK Gutschrift	ADA Cosmetics GmbH	10887160	2142.60	2549.69	
01	EK Rechnung	Abbyy Europe GmbH	2801488			99.0
01	EK Rechnung	Abbyy Europe GmbH	2801488	264.00	314.16	99.0
01	EK Rechnung	Richter Aluminium GmbH	47480			11.0
01	EK Rechnung	Richter Aluminium GmbH	47480			11.0
01	EK Rechnung	Richter Aluminium GmbH	47480			11.0
01	EK Rechnung	Richter Aluminium GmbH	225615	35690.71	42471.94	11.0
01	EK Lieferschein	Alde Gott eG	115640			

Figure 8: Check | change field values

The header data is displayed in the top area, and the position data is displayed in the bottom area (5). The changes can be made directly by clicking on the fields (1). Click Save (2) to accept the changed data. The fields that can be changed are specified. The entire record is removed by pressing Delete (3). Incorrect or unrecognized values are portrayed red. Notes provide information why this is so (4). In our case, the plausibility check revealed that net + VAT tax \neq gross, and therefore one (or more) values were incorrectly recognized. The same procedure can be used for the position data in the bottom area of the mask on the right (5).

The field values that are selected in the hit list are selected in the document display, and this allows you to quickly see where the value was identified and if it is correct.

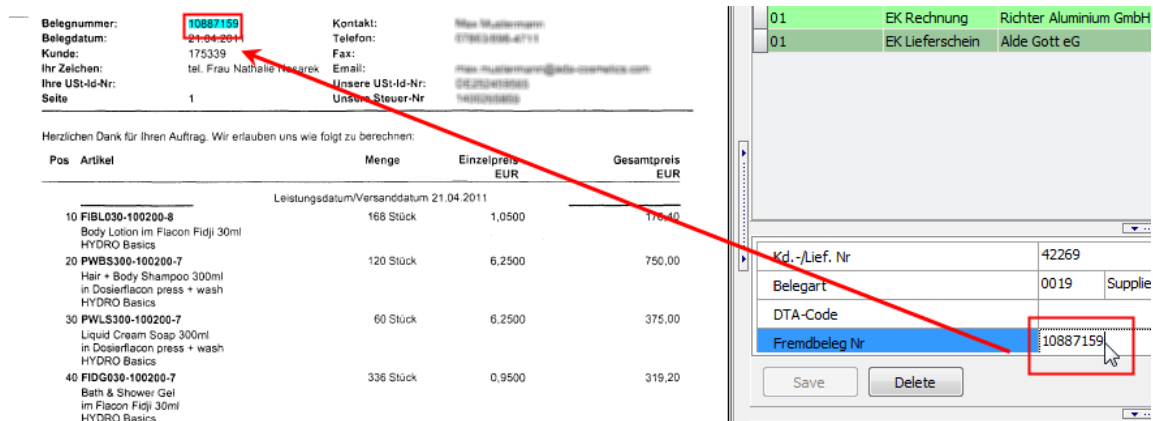


Figure 9: Display selected values

6. The recognition of the creditor/debtor

The creditor/debtor can be recognized in different ways. For each customer system, features (so-called identifiers) are saved that the document reader can use to identify the creditor/debtor. Identifiers are for example e-mail addresses, URLs, customer numbers, tax IDs, IBAN, account numbers, etc. Values are saved for these identifiers that identify each creditor/debtor. The document reader reads the raw text of the document and compares it with the saved identifiers. If the identifier is recognized, the system searches directly adjacent to the term (clockwise: right, down, left, up) for the associated value, and then it is assigned to the creditor/debtor. In addition, the saved templates are checked to see whether the values can be assigned to the creditor/debtor according to the identification features saved in the templates.

Training identifiers while forming dossiers

A series of terms that can be used as identifiers are already saved in the system. However, there are always different spellings or words that can be used as identifiers. These deviations can be saved in the system, and the document reader is trained. After the document reader is trained, the values can be correctly identified in the future, and less manual intervention is required. Remember: When the identifiers are saved, they apply to all creditors/debtors. This is particularly problematic in the case of terms that are not unique such as the date and document date. With some creditors/debtors, the date on the document is always the document date; in the case of other creditors/debtors, the date can be different, such as the delivery date.

It is a good idea to expand the identifier for the term *account*, for example, since it can be used in various ways: Account number, account No., acct. No., account. If this account number is saved as an identifying feature (identifier) for the creditors/debtors, the different possible terms need to be saved. As mentioned, this only needs to be done for a single creditor/debtor, and then it applies for all of them. (The commands can be found in the context menu of the document display).

str. 49, 80687 München
ien - Konto-Nr. 6255504
ünchen - Konto-Nr. 666
59-0 - Fa. Konto-Nr. 39 51

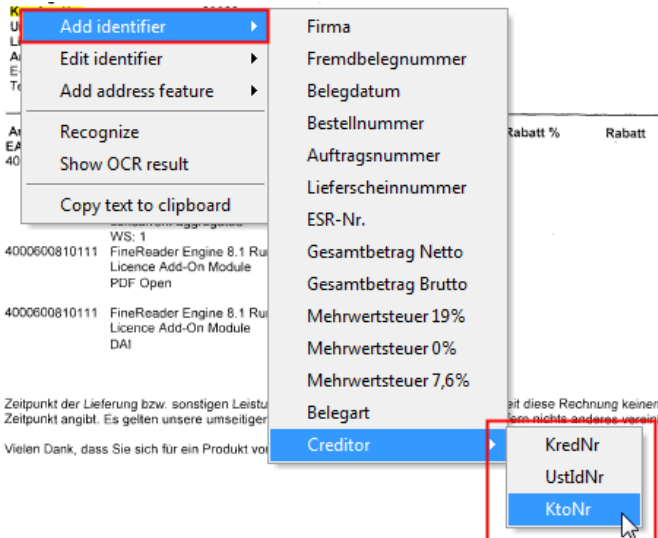


Figure 10: Adding identifiers (pop-up menu)

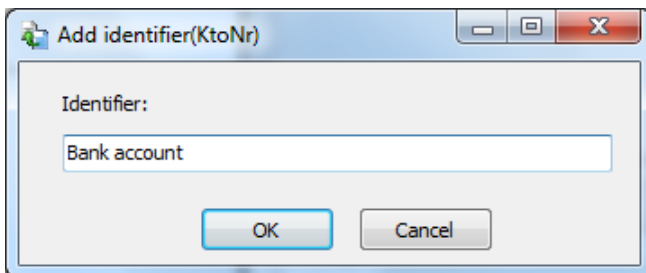


Figure 11: Defining the identifier

The recognized value can be accepted or entered manually. If other values need to be transferred from the document, this can be done by holding down the SHIFT key and clicking on the value.

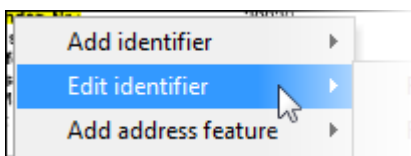


Figure 12: Menu selection

Use the command *Edit identifier* in the pop-up menu of the document display to display the saved values and delete them if desired.

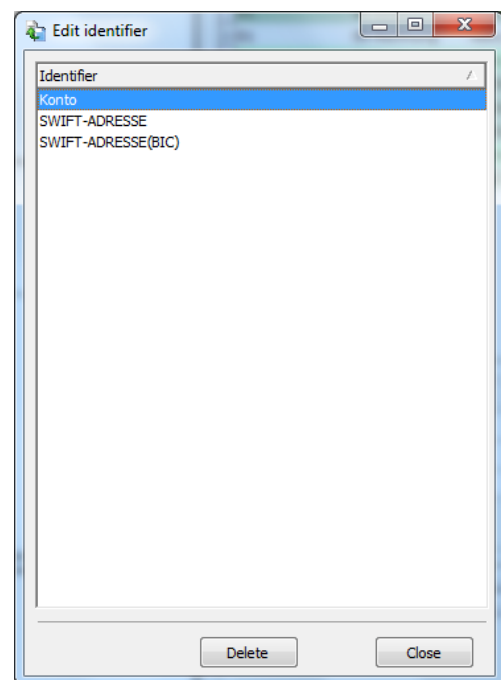


Figure 13: Editing identifiers

If for example a document type ID is desired, it also needs to be trained. The documents are scanned using a dummy document type and then recognized as the corresponding document type using specific identifiers.

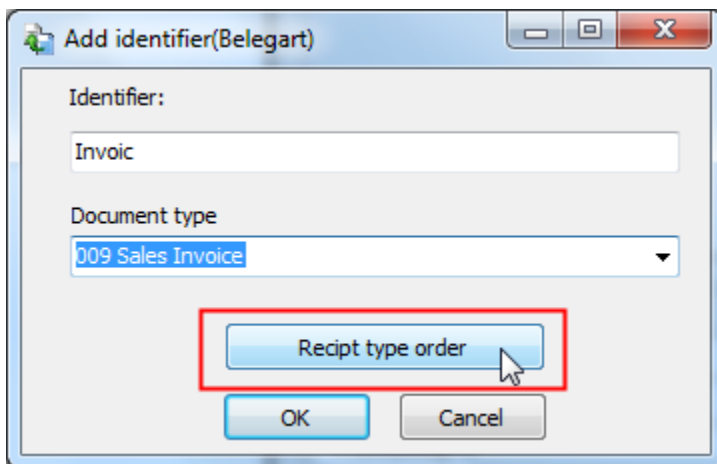


Figure 14: Adding identifiers

First the identifier is specified for the document type and the associated document type is defined.

The sequence of document types specifies the types of documents that can be associated with this identifier. For example, the expression "credit memo" is on credit memos, and frequently so is the word "invoice". If the expression "credit memo" is at the top, it is a credit memo. This is defined by the document type sequence (by simply moving the document types).

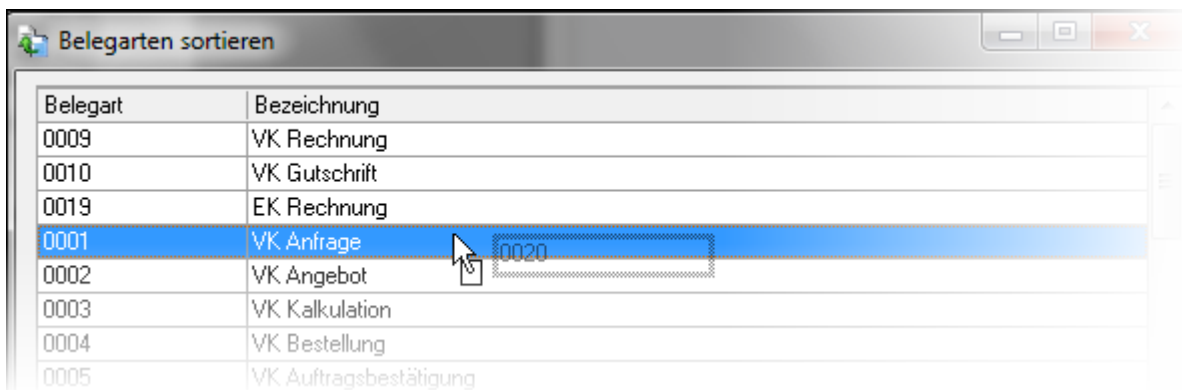


Figure 15: Sorting document types

Adding address features

To identify the creditor/debtor, the identifier is exported from the raw text along with the associated value directly next to it. Features are added for the creditors/debtors. If the creditor/debtor is recognized, the fields that need to be filled in are completed using the data from the address master. If the creditor/debtor is not recognized (field not filled in), data need to be added to the master data in the ERP program. The data from the ERP program are transferred to the HABEL address master and are available to be used in HABEL. The data saved in the document reader are not available for the ERP program.

Right-click the document to open the pop-up menu. Here you can enter address features. Ideally you will right-click a recognized value that is significant for this creditor/debtor. In this case, it is the account number **(1)**.

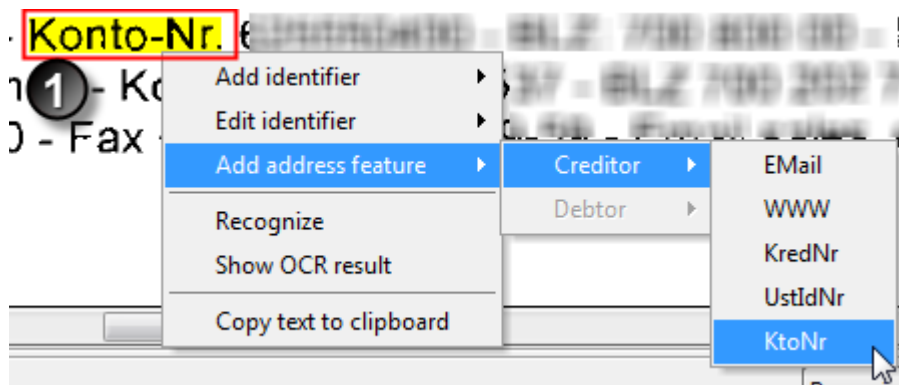


Figure 16: Address features in the pop-up menu

The saved terms need to be individually defined for each customer system and can therefore deviate from your system. After a feature is selected, the input mask appears.

Document reader
Document entry/recognition

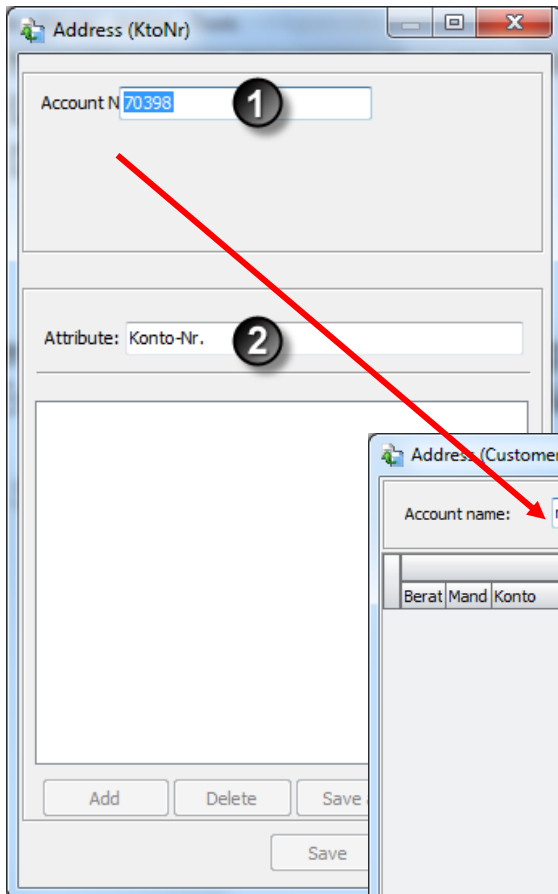


Figure 17: Editing address feature

The name of the creditor/debtor is entered in the account No. field **(1)**. The recognized field has already been transferred to the *Feature* field **(2)**. In the master access, use the tab key to navigate to the data saved in the address master.

Click the *Search* button **(3)** to check whether any master data exist. If they exist, they can be transferred. The recognition of the creditor/debtor is expanded to the newly saved feature, if it was added as such.

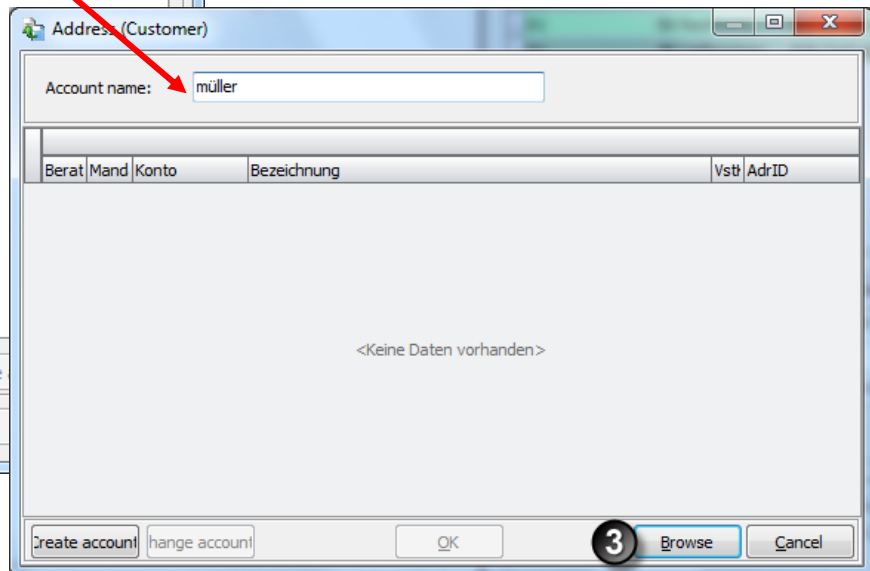
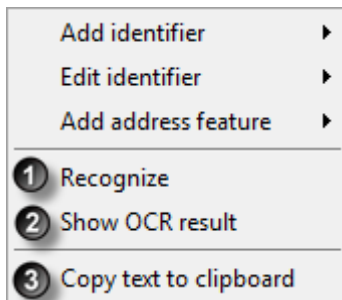


Figure 18: Master data access

7. Other commands in the program "Forming dossiers"

Pop-up menu of document display



The recognition of this document can be restarted by selecting the command *Recognize* (1). This is used for example when new features are added for recognizing the creditor/debtor. The data are added during the new recognition.

By selecting *Copy text to the clipboard* (2), the recognized value is available to be inserted at the desired location.

Figure 19: Pop-up menu (additional commands)

Select the command *Display OCR result* (3) to display the recognized values. They are grouped in the freeform result, template (according to pages) and header/position data. Note: For the results to be displayed, access the command and start a new recognition by clicking the *Recognition* button.

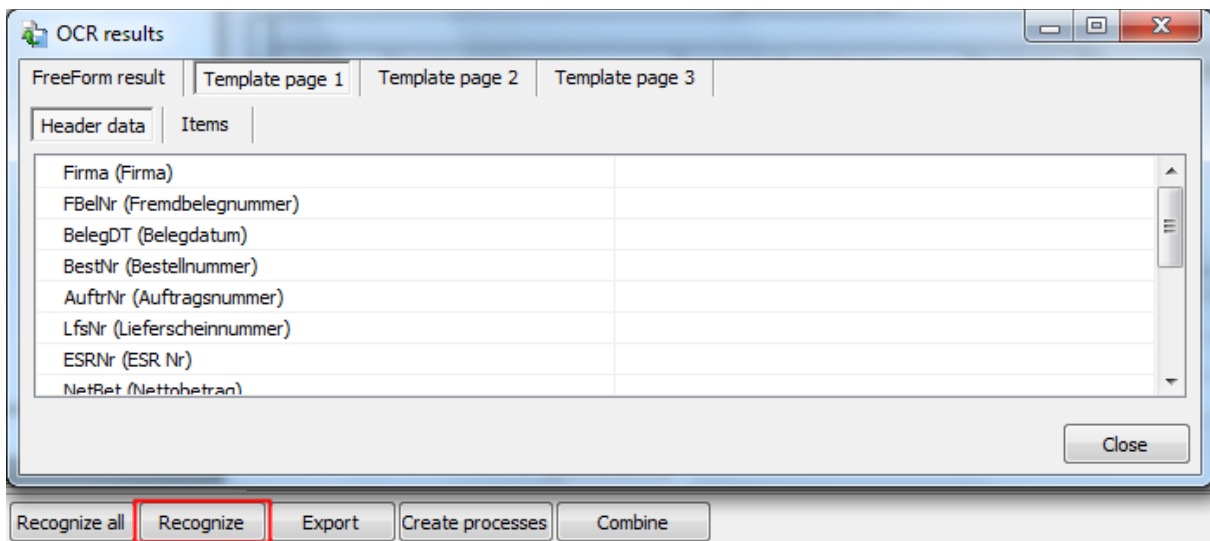


Figure 20: Display OCR results

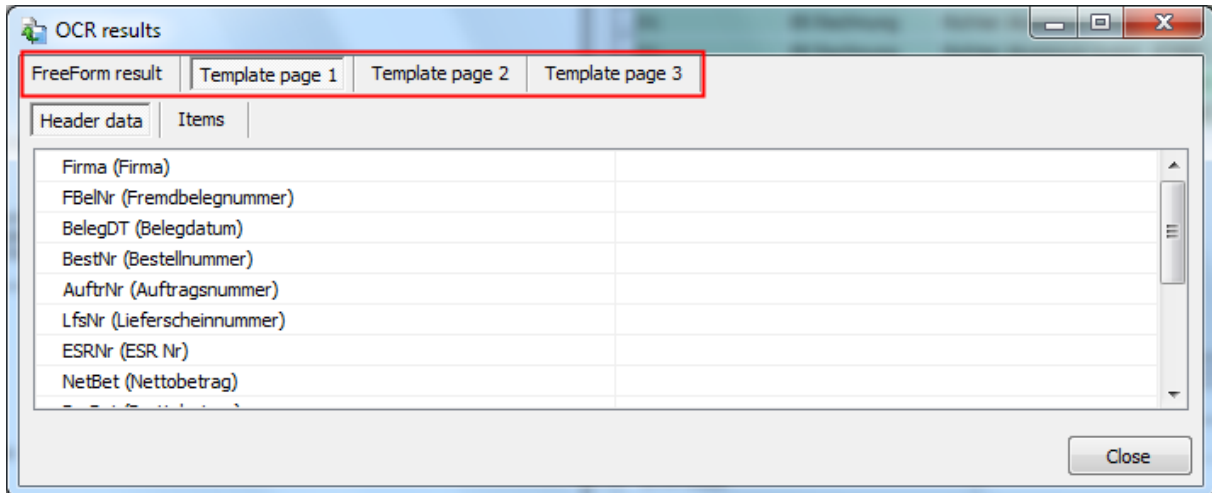
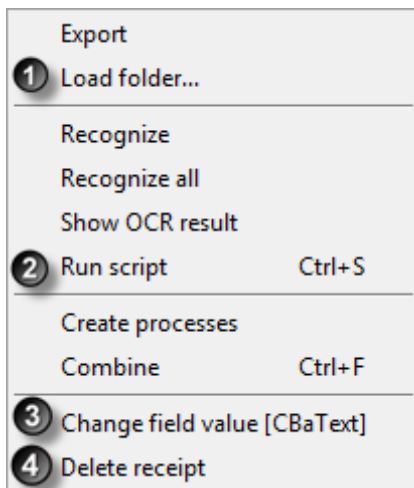


Figure 21: Display OCR results

An overview is displayed that shows which data that were recognized by freeform and which were recognized by the template (organized according to template pages). The listed data are raw data, that is, even if data manipulation was saved, the actually recognized values are displayed.

If errors occur during recognition, the overview can help find a solution since it displays what actually was recognized.

Pop-up menu in the hit list of recognized values



The commands *Export*, *Recognize*, *Recognize all*, *Form dossiers* and *Join* are also available as buttons in the footer of the mask. Their use has been described above. In the previous section, it was also explained which data can be retrieved with the command *Display OCR results*.

In addition, the following commands can be accessed in the context window:

- Load folder
- Run script
- Change field value
- Delete document

Figure 22: Pop-up menu

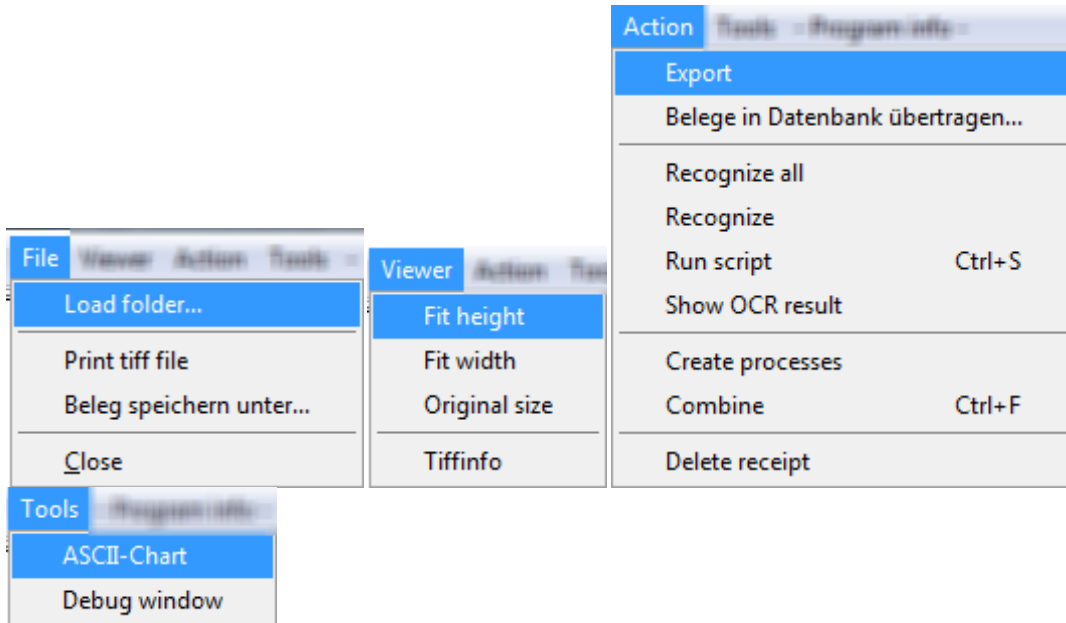
Select the command *Load folder* (1) to retrieve the folder selection and process additional stacks in the program *Form dossiers*.

If scripts are saved such as the plausibility checks of the values or separation logic, they can be started by the command *Run script* (2). If the command is not executed and then *Form dossiers* is run, the scripts are run at this time.

Select *Change field value* (3) to change the field that was right-clicked. An input mask appears in which the desired value can be entered directly.

Select the command *Delete document* (4) to delete the selected document. Remember: The document is also not in the archive.

Menu bar



Various known commands are available in the menu bar. Additional commands will be briefly explained below:

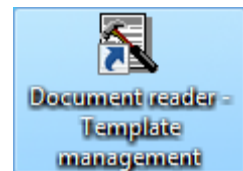
- Print tiff file: To print the document
- Save the document under: Can be used to create templates
- Viewer: Adapt the document display and information to the tiff file
- ASCII chart: Display the characters that can be used in ASCII format
- Debug window: Further details when creating the script

Create templates

General

Templates are also created to optimize the readout of document information. Many business transactions are printed on documents that have an identical structure. Document information such as the document date, document number, article data and prices are always found in the same location. These locations can be defined in a template and provided to the document reader so that the data can be recognized faster and more accurately.

The program hdhab404 is available to create the template.



Explanation of terms

The regular terms associated with the document reader will be used in the descriptions to follow and will be briefly explained below:

Template

Is the term for the set layout that can be filled with content. In our case, it is a framework for defining the locations where values can be read out.

Freeform

With freeform recognition, the raw text is read out and compared with terms that are defined to identify values (such as creditor, invoice number, etc.). If the searched term is found, a search is performed directly next to it (clockwise: right, bottom, left, top) for the value that is then used as the read-out information.

Identification

The document reader uses the identification to identify the associated creditor/debtor. An indicator is saved with a fixed coordinate which is used to determine the creditor/debtor.

Setting an anchor

An anchor is placed at the location where a certain term can be found in the document and can hence be identified. Additional information about the entire document is selected and exported using this anchor. Coordinates are used to define the location at which an anchor is located and where the other data can be found in the document.

Short overview

The following offers a brief explanation of how to create a template.

1st step *SCAN THE DOCUMENT*

Use the KOFAX (300 dpi) scanning program, and select the document type to be processed by the document reader.

2nd step *SAVE THE DOCUMENT*

Save the document in tiff format using the *Recognize documents* program (hphab400). Note: The first, subsequent and last pages are saved as individual files. By the way: The documents can also be saved directly as a template using the *Form templates* program (hphab401).

3rd step *CREATE A NAME FOR THE TEMPLATE*

In template administration (hphab 404), click the New button, and enter the name of the creditor or debtor in the Name field to access the master data. Create a new record by clicking Save.

4th step *ASSIGN THE DOCUMENT AS THE FIRST PAGE OF THE TEMPLATE*

Select the new record and the term Identification, and re-click the New button. The associated tiff file is selected for the first page of the template which you can save by clicking Save.

5th step *IDENTIFY THE SUPPLIER*

By pressing the button New after selecting First page below the Identification, you specify the criterion/criteria by means of which the creditor/debtor is identified to use this template. Certain points must be observed which are described in detail under section 8 of this chapter.

6th step *DEFINE THE FIELDS/COLUMNS TO BE READ OUT*

After selecting the term Document, click the button New to assign the form to the first page. The anchor is saved, and then the positions are defined from which values are exported. Certain points must be observed which are described in detail under section 9 of this chapter.



If the anchor is changed, always use the test icon to see if the anchor is still correctly recognized.

7th step *SAVE ADDITIONAL PAGES*

Steps 4 to 6 are repeated to save the following pages as a template.

8. Scan the document (first step)

In the KOFAX scanning program, the document is scanned as the document type that will be used for processing in the document reader. Check the following:

- Is the document a single page?
- What do the second, third and other pages of the document look like?
- Are there foreign language versions of the document?

To optimize recognition, make sure that the scanning optimization software KOFAX scans the document with 300 dpi.

9. Save the document (second step)

In the *Recognize documents* program (hphab400) or Form dossiers (*hphab401*), each page of the template is saved as a single TIFF file.

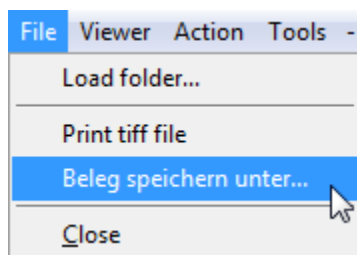


Figure 24: Saving procedure

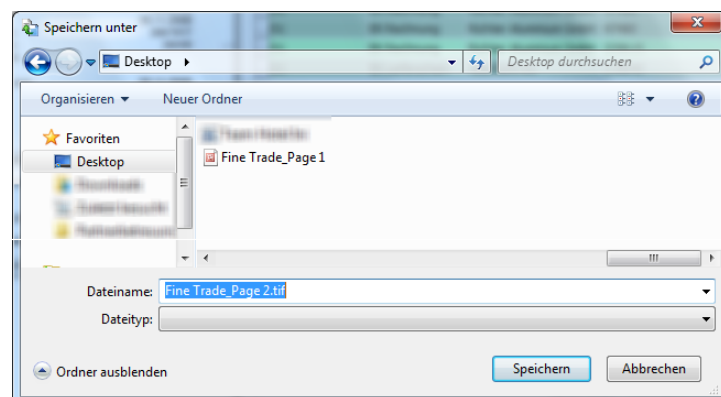
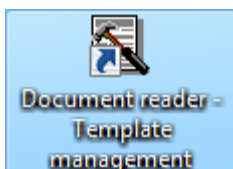


Figure 23: Menu selection

Use the program hdnab404 for the next steps.



10. Create a name for the template (third step)

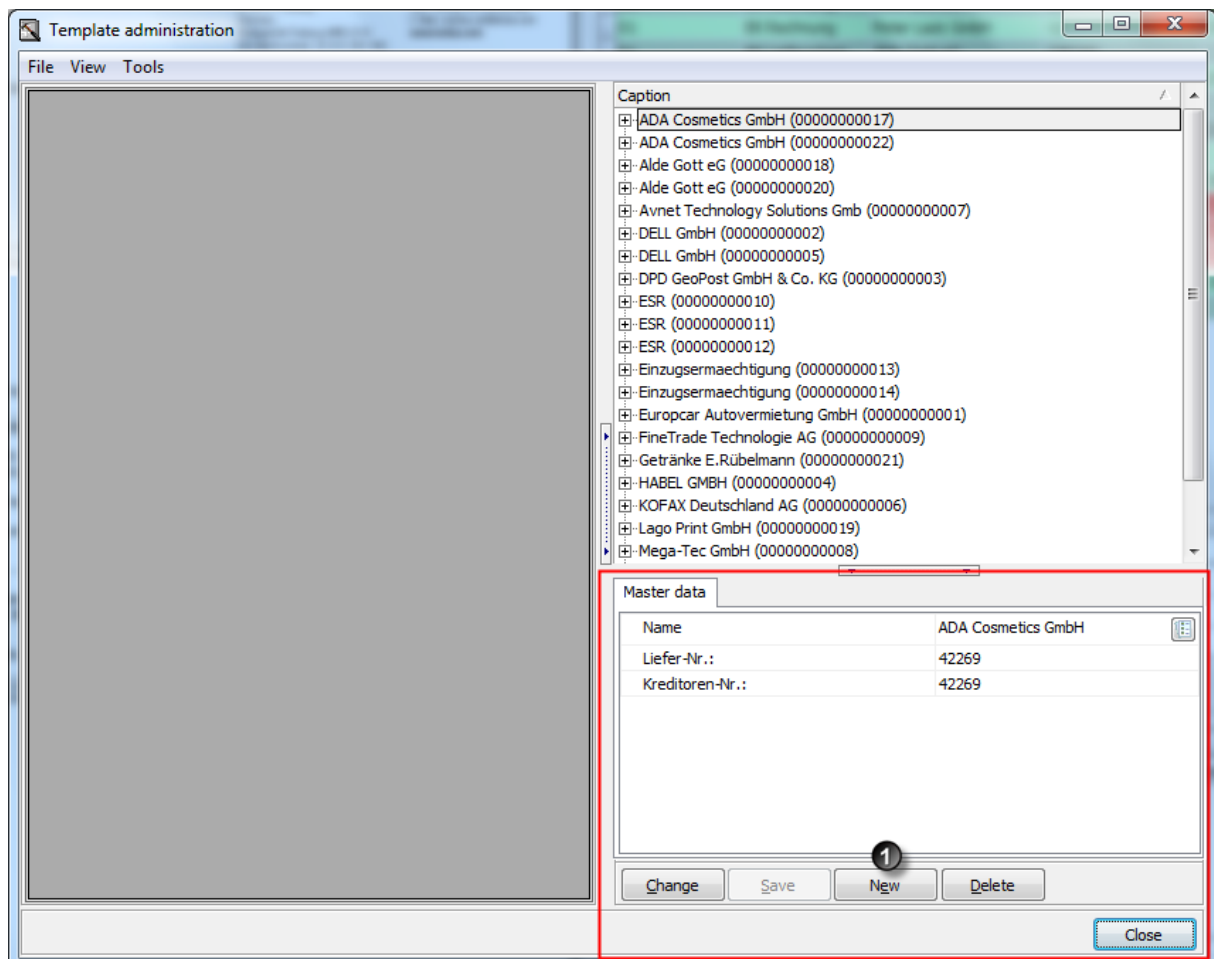
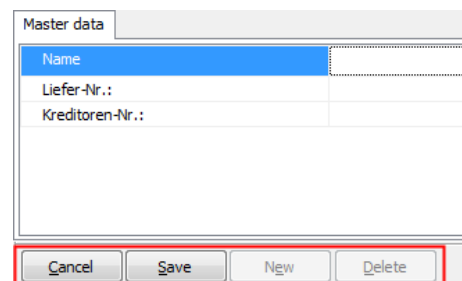


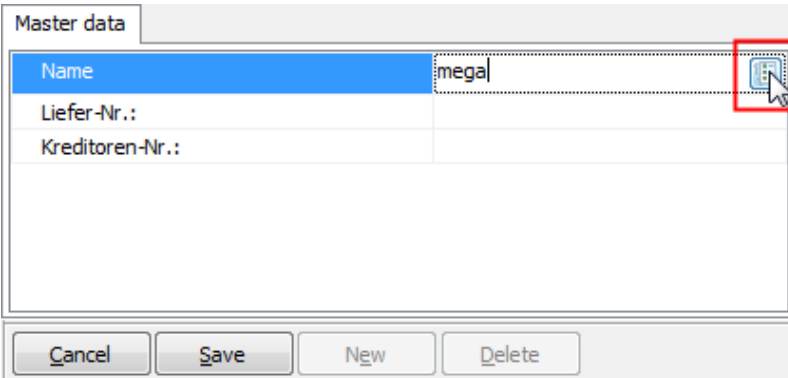
Figure 25: Start mask for template administration

In Template administration, click the *New* button (1). In the bottom section of the mask, the fields are emptied, and the buttons *Cancel*, *Save* and *Delete* are active.



Document reader
Create templates

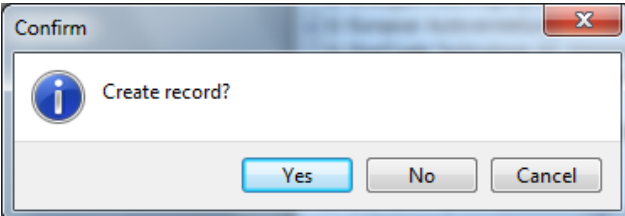
In the field Name, enter the name of the creditor/debtor, and click the icon at the end of the field. A selection box appears with the created address data. After you accept the data, transfer the data by clicking the OK button. You can change the name (that is, the designation of the template) later. If you have saved a template for a supplier which can be used for another supplier, name this template using the conventions of, for example, SAP templates.



The screenshot shows a 'Master data' form with the following fields: 'Name' (containing 'mega'), 'Liefer-Nr.:', and 'Kreditoren-Nr.:'. A selection icon is visible at the end of the 'Name' field, highlighted with a red box. At the bottom of the form are buttons for 'Cancel', 'Save', 'New', and 'Delete'.

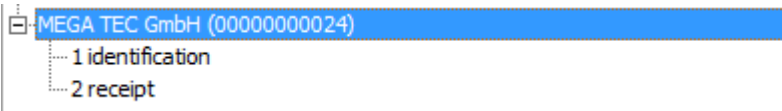
Figure 26: Accessing master data

The number of the supplier/creditor is added. Save the data by clicking the Save button, and confirm the record by selecting Yes. The data can be seen in the tree structure.



The screenshot shows a 'Confirm' dialog box with the text 'Create record?' and an information icon. Below the text are three buttons: 'Yes', 'No', and 'Cancel'.

Figure 27: Confirmation request

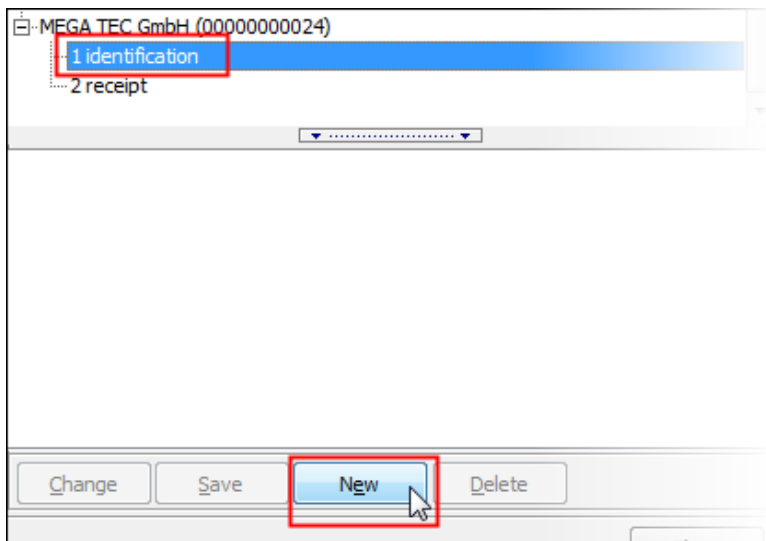


The screenshot shows a tree structure with the following items: 'MEGA TEC GmbH (00000000024)', '1 identification', and '2 receipt'.

Figure 28: Assigning the supplier

11. Assign the document as the first page of the template (fourth step)

Add the entry of the new template, then select the term Identification. The bottom section of the mask is emptied, and no master data are shown. The button New is active; click it.



In the Identification mask, define the page that is to be saved **(1)**, and upload the associated form **(2)**. Create a record by clicking Save.

Figure 30: Identifying documents

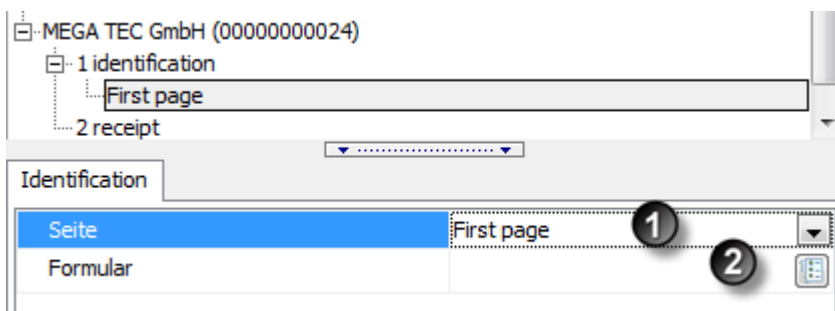


Figure 29: Defining the *first page*

Document reader
Create templates

12. Identify the supplier (fifth step)

To identify the first page of the template, you need to specify the locations where terms are to be exported. After selecting the term First Page below the identification, click the button **New**.

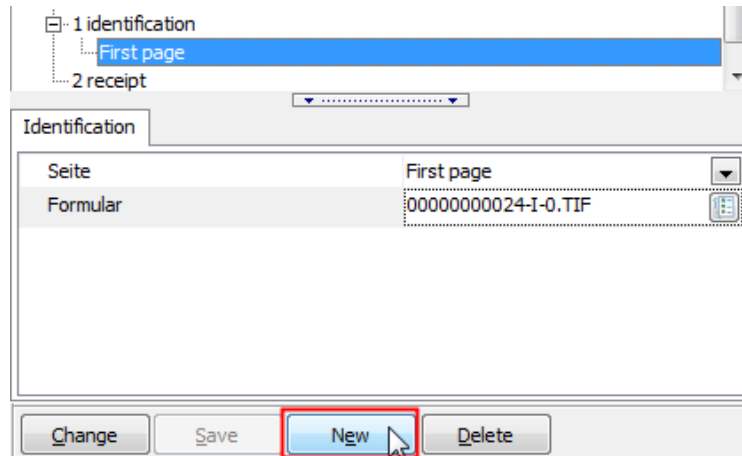


Figure 31: Identify the supplier

The Identification tab appears with the two criteria *% match* and *Attribute*. Use *% match* to set the percentage with which the recognized value must match the specified *Attribute*. For example, a 70% match for a term consisting of 10 letters means that 7 letters need to manage.

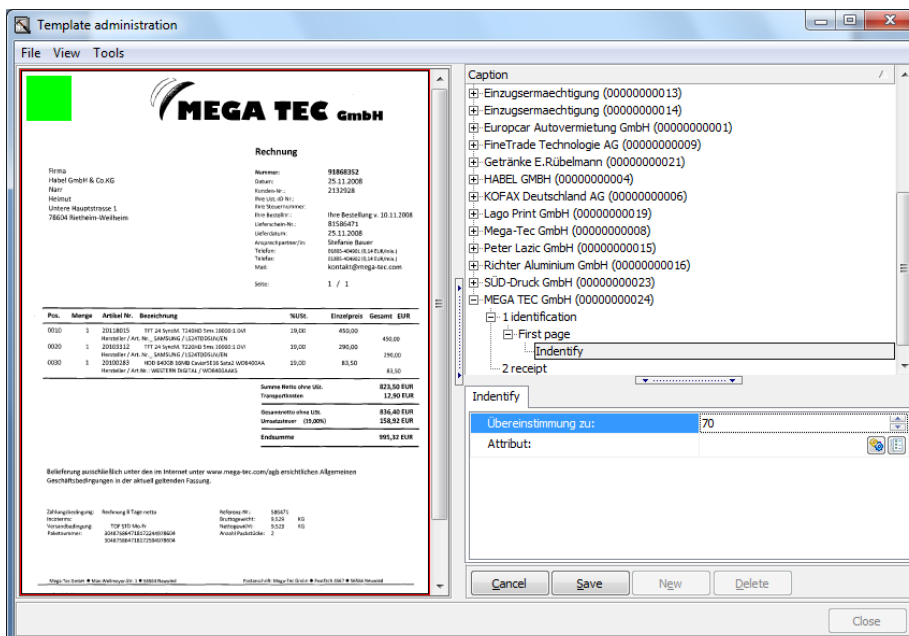


Figure 32: Specify % match



The document used for saving the template is a sample. When defining the criteria/attributes for identification, it should be noted that the incoming documents may deviate (skewed text, poor quality, etc.). The green area specifies where the attribute is located for identifying the supplier. The area should therefore not be too close around the section within which the attribute is located.

To save an attribute, a section and the text to be searched need to be defined. Click the icon and enter the text that is within the green area.

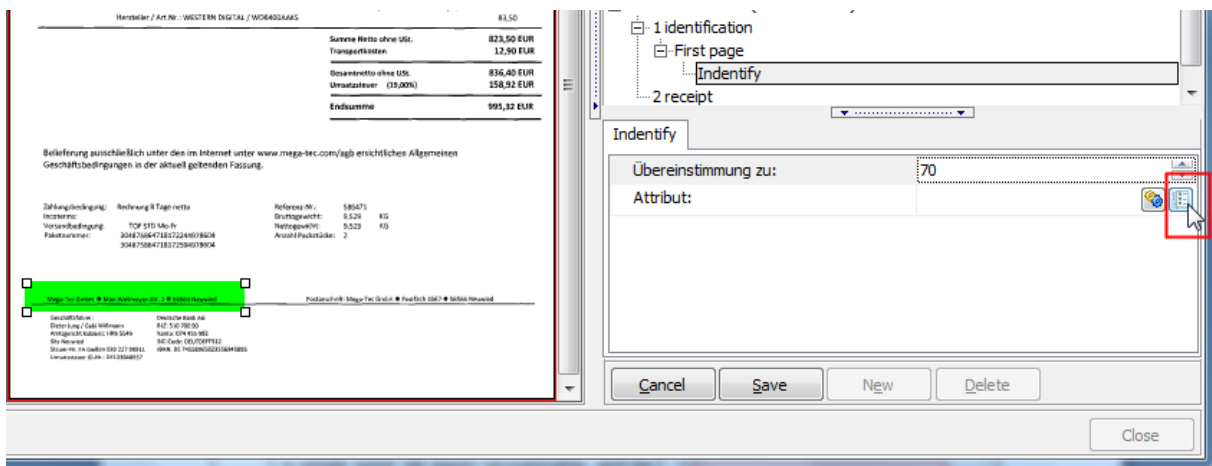


Figure 33: Define the text area for the attribute

Several texts can be saved in the attribute list that are used as an attribute for identification. They are separated with a vertical line (|) in an attribute field.



Suitable terms for attributes are those that are always at this location on the suppliers form (the section defined with the green area) and therefore clearly identify the supplier. The attribute can consist of elements of the letterhead or a printed term. It is important for the attribute to always be located at the same place, consist of a single word, be easily recognizable, and be unique. Several terms that appear together can be used for the identification.

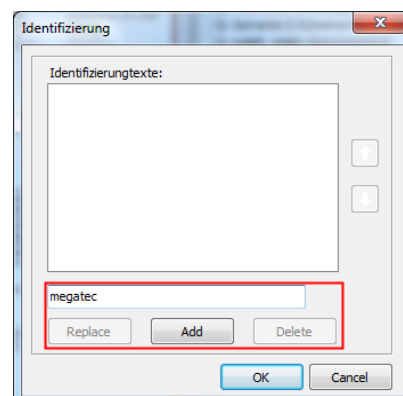


Figure 34: Attribute list

Click the test icon to check if the document reader can find the attribute in the green section.

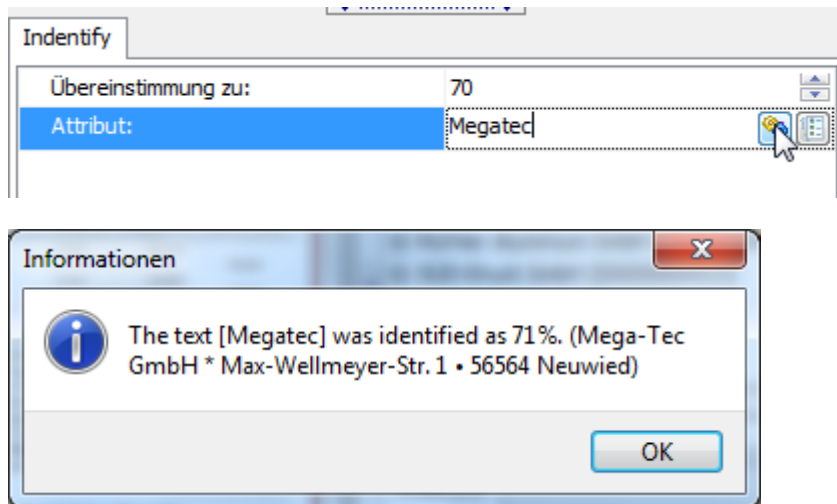



Figure 35: Checking the recognition of the attribute text

Please note: These attributes should be valid only for this supplier to prevent wrong templates from being used.

	<p>When saving terms, be sure to note which letters are uppercase and lowercase. If there are several templates for a single supplier, the definition of the identification attributes must contain special distinctive features.</p>
---	---

Below are a few examples of identification features that can be used.

- Company name from the letterhead
- Company address (street address and city)
- Website address from the header or footer

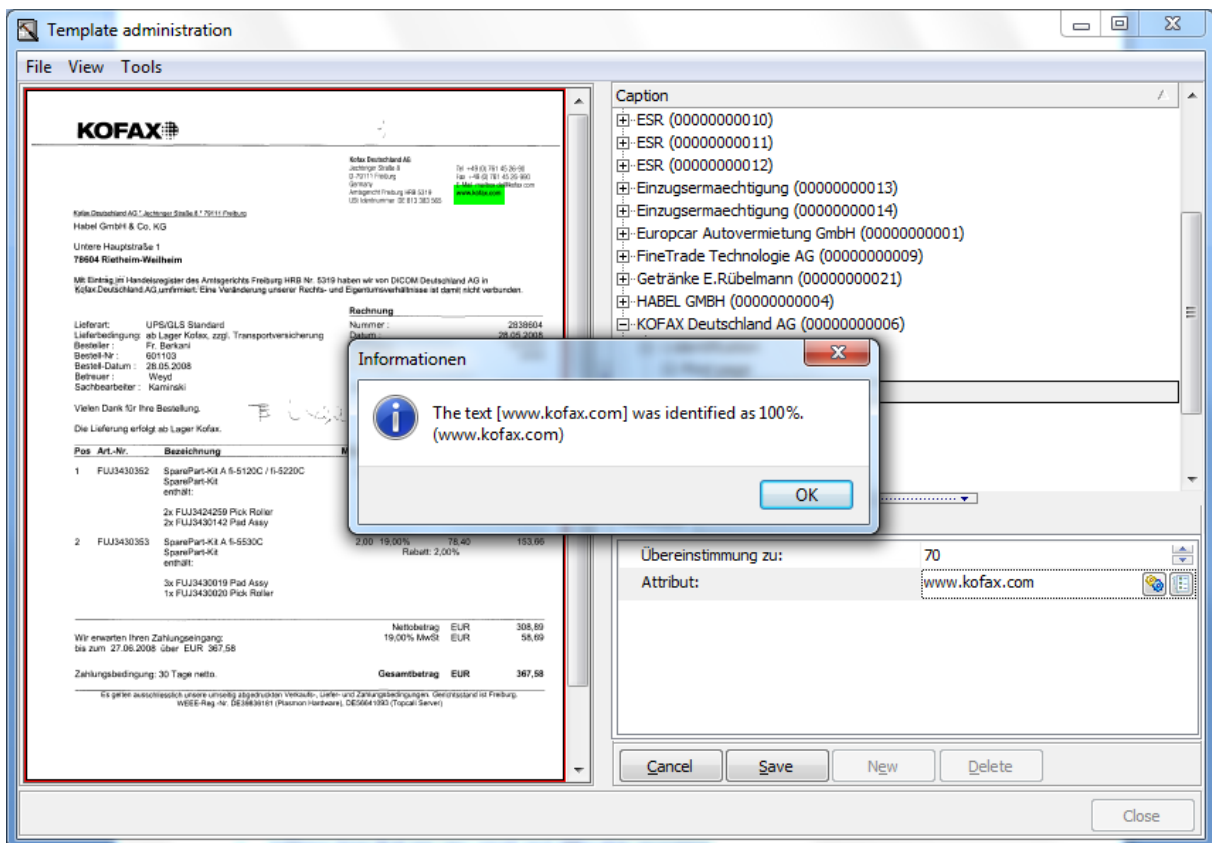
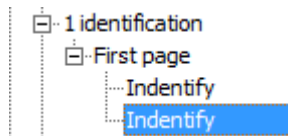


Figure 36: Identification using the Internet address

If several attributes are saved, all of the attributes must be recognized for the identification (depending on the saved % match).



Identification features that cannot be used

- Logos that have icons and no clear text
- Address lines that are very small and difficult to recognize

13. Define the fields/columns to be read out (sixth step)

To define the fields and columns from which data are to be exported, select 2-Document under the template name.

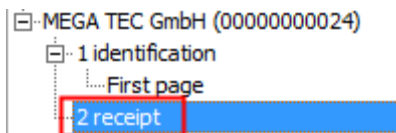
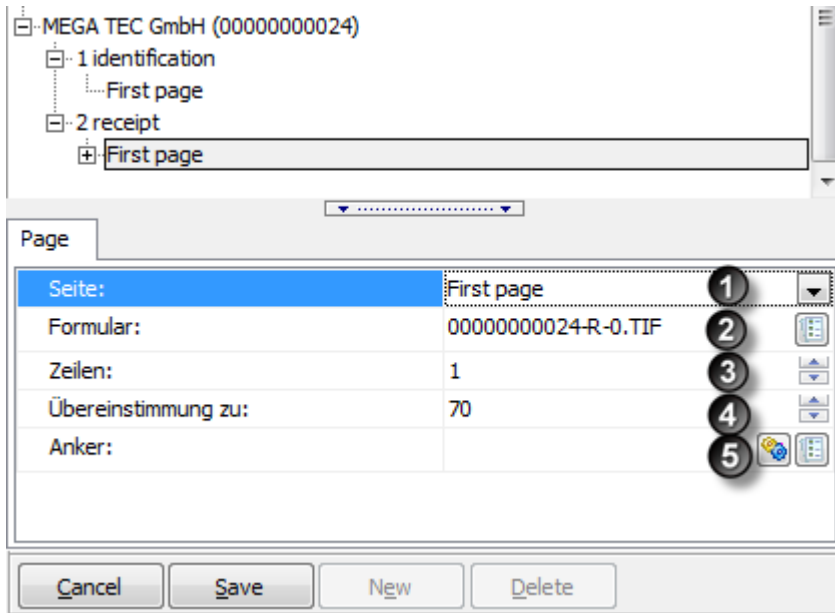


Figure 37: Defining the fields and columns

Click the *New* button in the bottom half of the mask to activate the input mask for the document.



Select the relevant page **(1)**, and upload the associated form **(2)**. If the form has positions that always consist of the same number of lines, the number of these lines can be saved under the criterion **(3)**. If the number of lines differ, enter a 1. In addition, an anchor **(5)** is set to identify the first page of the document, and the % matches **(4)** are saved.

Figure 38: Document - First page

Document reader
Create templates

The anchor is the attribute that identifies the document. The additional information is selected and exported starting from this location. The turquoise field is set at the location at which the same term always occurs and identifies the document (in our case, the term "invoice"). In the anchor field, the text to be identified is defined by the anchor text button. Use the test icon to check the degree to which the term is recognized (taking into account uppercase and lowercase). When selecting the anchor, remember that it is a term that is printed. If the printing is skewed in the standard layout, the other passages to be exported are still selected and recognized starting at the anchor. Only one anchor can be set for the document.

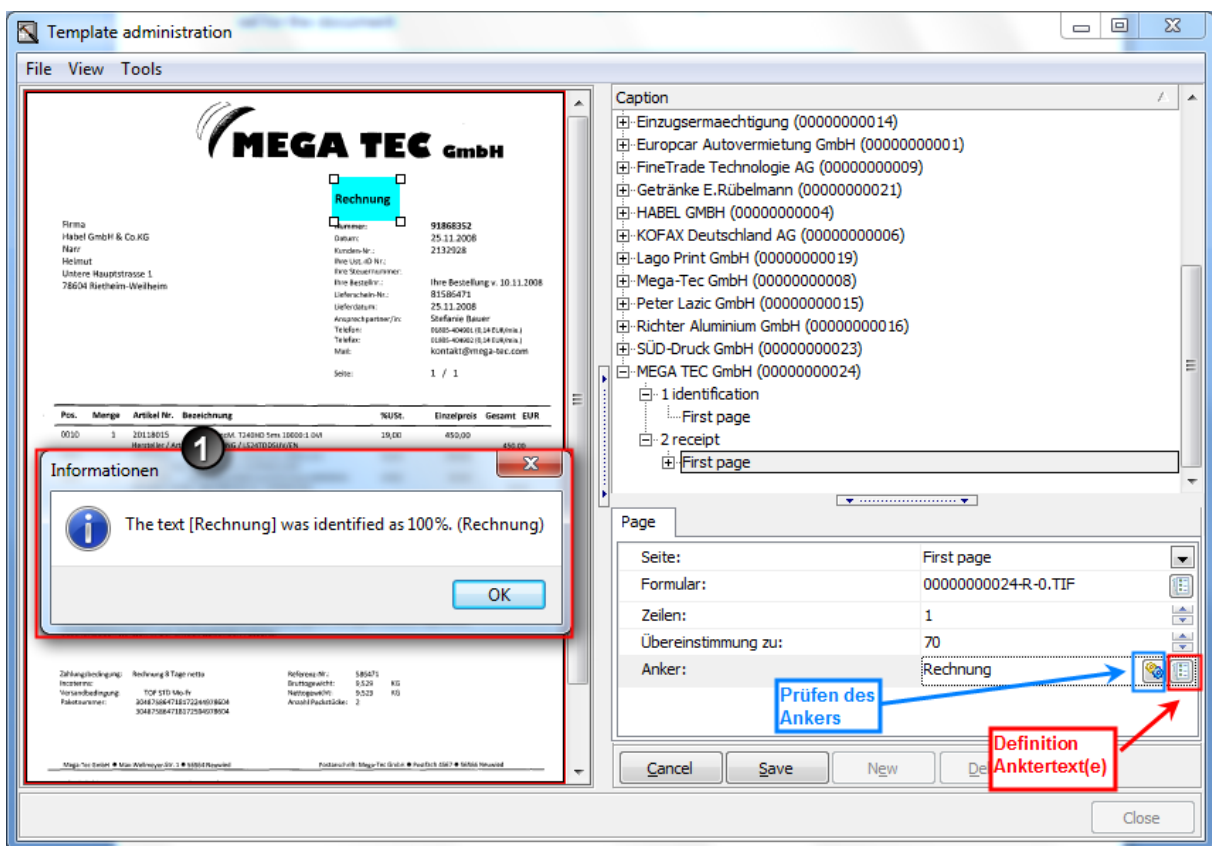


Figure 39: Recognizing the anchor text

	<p>The recognition of this anchor is very important since the placement of the other values to be recognized depends on it. If the anchor is changed, run recognition again (test icon).</p>
--	--

Document reader
Create templates

Below are a few examples of what can be used as an anchor.

- Descriptions that can be **permanently** affixed to this location for this creditor/debtor

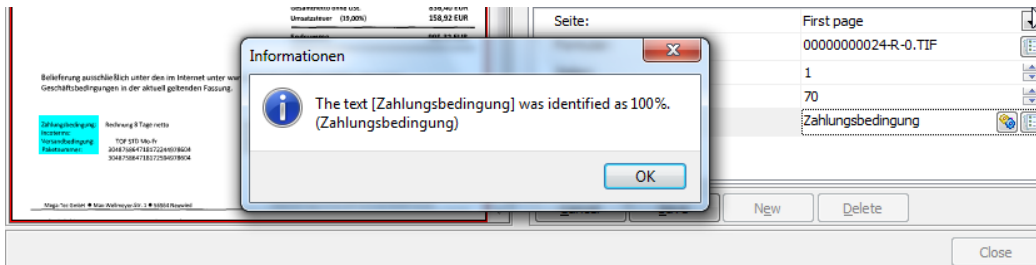


Figure 40: Examples of anchor text: Payment condition

- Address fields

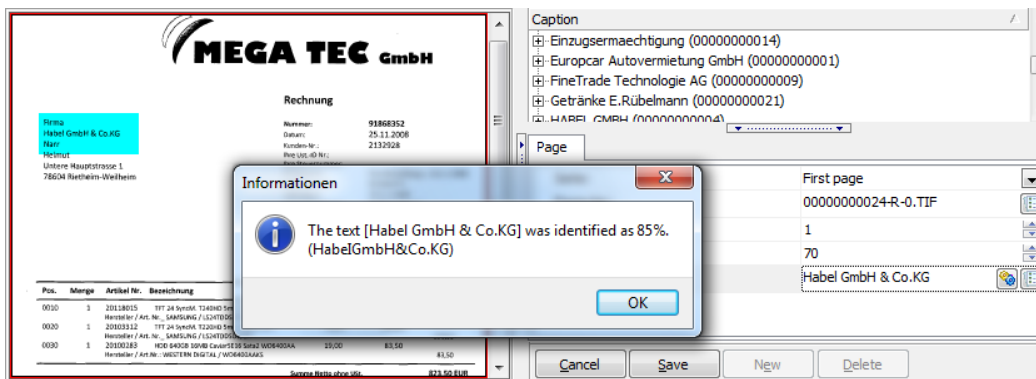


Figure 41: Examples of anchor text: Invoice address

- Invoice date or document date that can be assigned to this creditor.

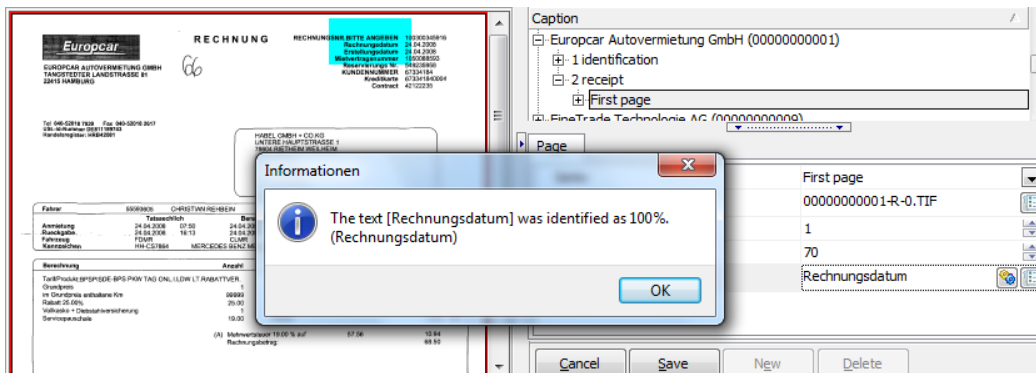


Figure 42: Examples of anchor text: Date

- Invoice or in this case the invoice number (example of deviating following page)

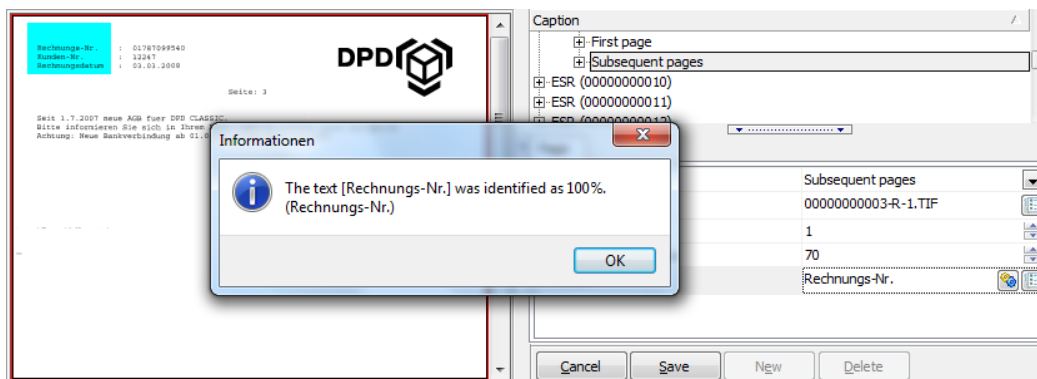


Figure 43: Examples of anchor text: Term: invoice

Terms that cannot be used as anchors

- Address field when several clients are saved
- Document passages that are part of form that, when the text shifts, prevents the other document information from being selected.

Using the anchor list, alternate terms can be defined for the same template with the same alignment. In our case, the term CREDIT MEMO could also be recognized instead of the term INVOICE. The document reader would, however, have to react the same in both cases (i.e., the additional document information is selected and exported based on the terms INVOICE and CREDIT MEMO). In the anchor field, the terms are separated by a | .

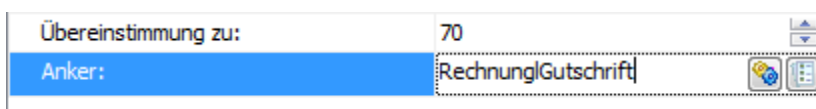
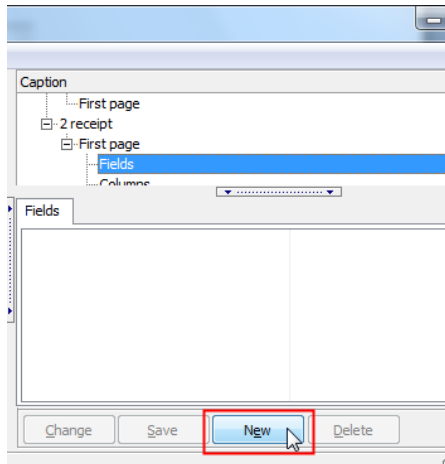


Figure 44: Several anchor texts

Document reader
Create templates



After the anchor is set, the fields and columns are defined from where the values are to be exported. After selecting the term Fields under 2-Document | First Page, click the *New* button.

The mask then allows you to enter information for the field to be exported. The yellow shading in the document display indicates the area where the information is located for the associated data field.

Figure 45: Definition of fields

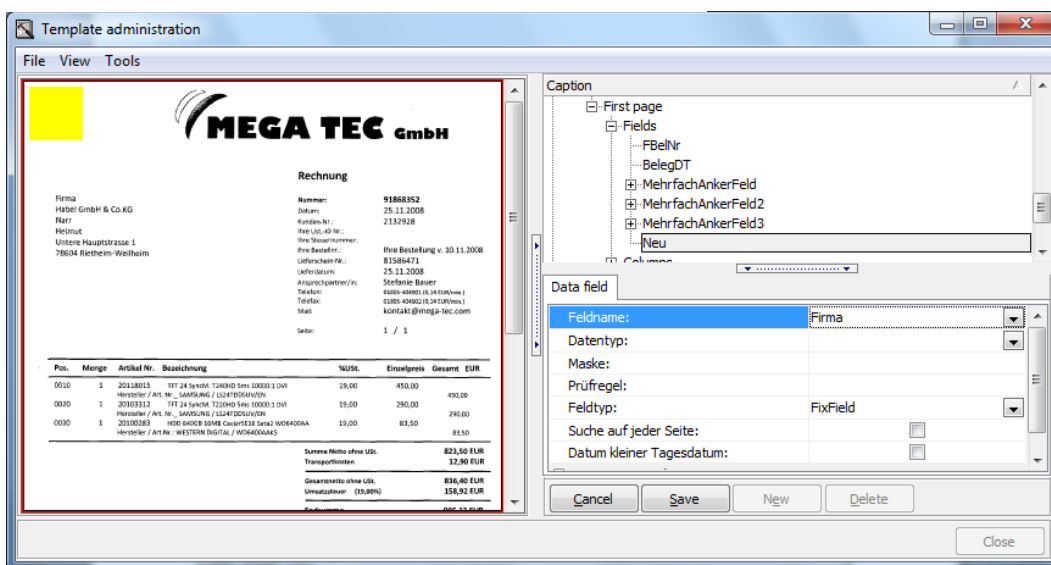


Figure 46: Definition of fields

Note that the box needs to be placed **closely** around the values **(1)**. If the area is too large, directly adjacent characters or values can be exported as well and lead to misidentification. Note: If the value to be exported could be longer than in the template, the size of the area needs to accommodate the potential maximum **(2)**.

Nummer :
Datum :
Auftragsnr.:



2838604
28.05.2008
2819523

Figure 48: Definition of box

Nummer :
Datum :
Auftragsnr.:



2838604
28.05.2008
2819523

Figure 47: Definition of box

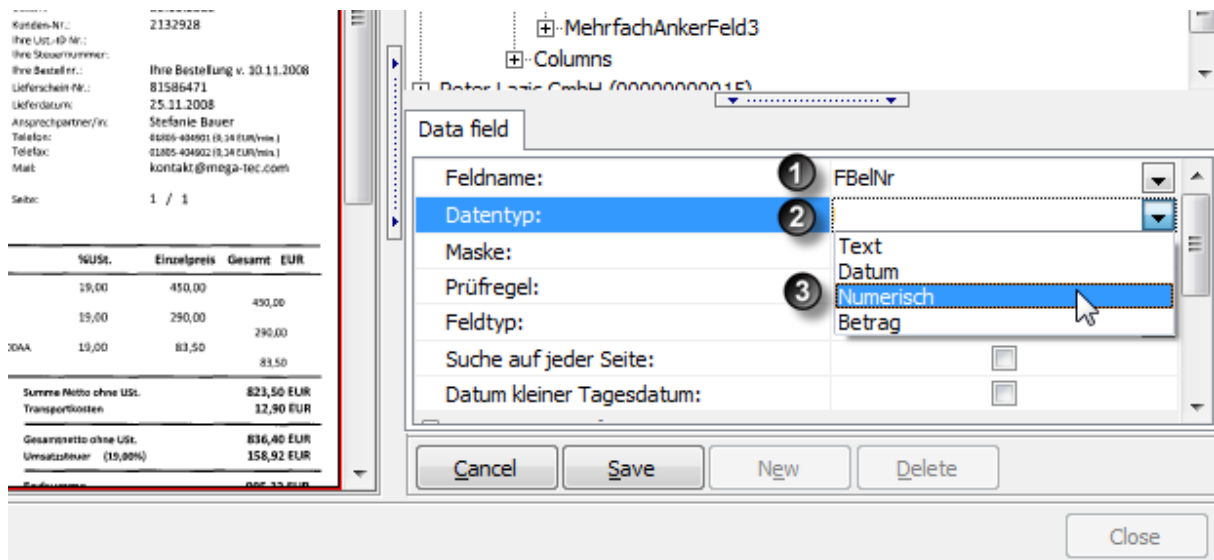


Figure 49: Definition of data type

Select the field name **(1)** from the list, and assign a data type **(2)**. Values are predefined that apply to your system. The validation rule **(3)** is automatically filled in, although it can be manually changed. The previously saved validation rule is retained even if the data type is changed.



To update the validation rule, press the keys CTRL + F5. The validation rules are based on regular expressions.

Below is an overview of regular expressions:

Characters	Meaning
.	(<i>period</i>) A period stands for any character.
[x-y]	This indicates a range of characters that are checked. Example: [1-3] only 1,2,3 are valid
[^x-y]	Negates the range of characters, that is, all characters with the exception of x-y are valid
\d	Decimal numbers
\D	All characters with the exception of decimal numbers
\s	Whitespace (spaces, tabs, etc.)
\S	All characters with the exception of whitespaces
\w	All characters found in words (letters, numbers, underscores)
\W	All non-word characters
\n	New line
\r	Carriage return
\\	\
\t	Tab
?	Checks whether the character does not occur or occurs once
+	Checks whether the character occurs at least once
*	Checks whether a character does not occur or whether it occurs any number of times
{x}	Checks whether a character occurs precisely x times
{x,}	Checks whether a character occurs a minimum of x times
{,x}	Checks whether a character occurs a maximum of x times
{x,y}	Checks whether a character occurs a minimum of x times and a maximum of y times
^	Checks for the beginning of the line
\$	Checks for the end of the line
\.	(<i>Backslash period</i>)Checks for a period
(xxx)	Describes a block (term) of regular expressions. Let us assume that you want to check whether two decimals occur just once in a chain of characters: \d{2} to check whether this condition occurs once or never, we need to place this block within parentheses (\d{2})?
	"Or" – this allows you to check several expressions, for example: (M m)? checks whether M or m occurs once or never.

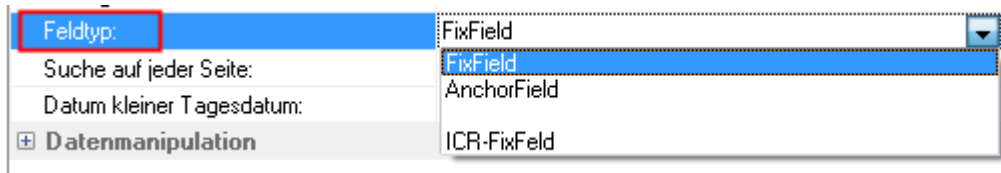


Figure 50: Definition of field type

There are three different field types that can be saved

FixField	The value is always at this location.
Anchor Field	The value is directly adjacent (to the right, left, above or below) this term, and its placement is defined by an area. To the entered data are added the anchor text, the % <i>match</i> and the position in relation to the found anchor text (left, right, top, bottom)
ICR FixField	The value to be exported is precisely at this location, but it is a handwritten value. Details can be set by adding information after selecting the ICR FixField in reference to the type of handwritten text and, if applicable, the number of available cells. Note: A separate ICR license is needed for ICR recognition (ABBYY).

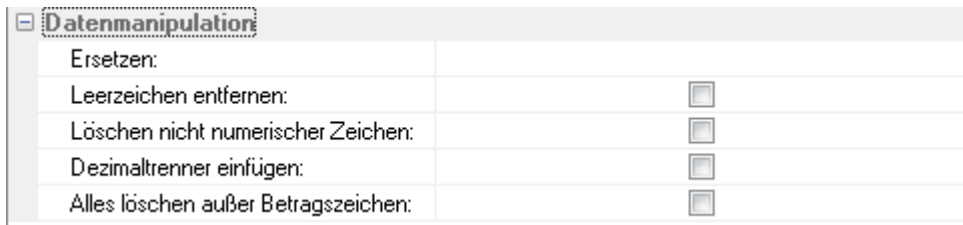
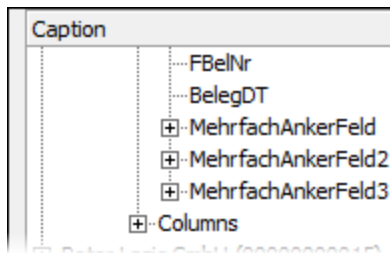


Figure 51: Data manipulation

The recognized values can be influenced by data manipulation

Replacing	If characters such as O and 0 or B and 8 are incorrectly recognized, they can be automatically replaced. Replacement rules can be used for any type of character (such as a period and space). The replacement rule is saved within square brackets and is separated by a , (comma) (alternately, ASCII values can also be used – see the menu <i>Tools – ASCII Chart</i> E.g.: [O,0] [B,8] [.,]
Remove space	Removes all spaces in the field content
Does not delete numerical characters	Removes all non-numeric characters with the exception of periods, minus signs and commas (.-,)
Insert decimal separator	Inserts a period two places from the right E.g. 2500 was recognized which becomes 25.00
Delete everything with the exception of the characters signifying an amount	Removes all non-numeric characters including periods, minus signs and commas (.-,)

Columns within which values are exported are created in a similar manner. Columns are primarily used for the recognition of item data such as article descriptions, article prices and quantities.



After selecting the term Column under 2-Document | First Page, activate the input mask by clicking the New button.

Figure 52: Define columns

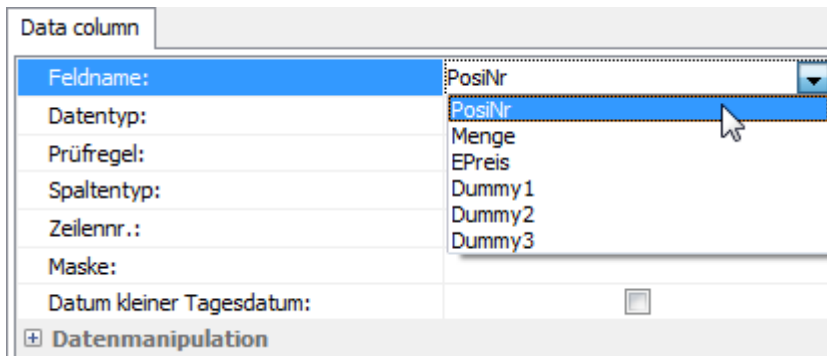


Figure 53: Define columns

Under *Field name*, select the database fields to which the recognized values are to be exported. The area within which the values are placed is marked yellow. Please remember that a model is used to set up the template that may deviate from the actual incoming documents. The area needs to be created with this in mind. In this case, the invoice could contain other article numbers that are below the listed numbers. The column therefore needs to cover the entire area within which the values can or could be located. *Data types*, *Verification rules* and *Data manipulations* can also be defined for the columns. When a multiple-line position is defined in the template, the *Line No.* can specify the line from which the value is to be exported.

14. Save additional pages (seventh step)

The cited steps need to be performed for each deviating following page.

15. Special case: Multiple anchor field

To export values that are at a specific location, fixed fields are created for exporting values that are directly adjacent to a specific term, that is, anchor fields. It is also possible to export values that are on the same line as an anchor, but are distant from the anchor. To accomplish this, multiple anchor fields can be defined (within the *Fields* definition **(1)**, select the field name **(2)** multiple anchor field **(3)**).

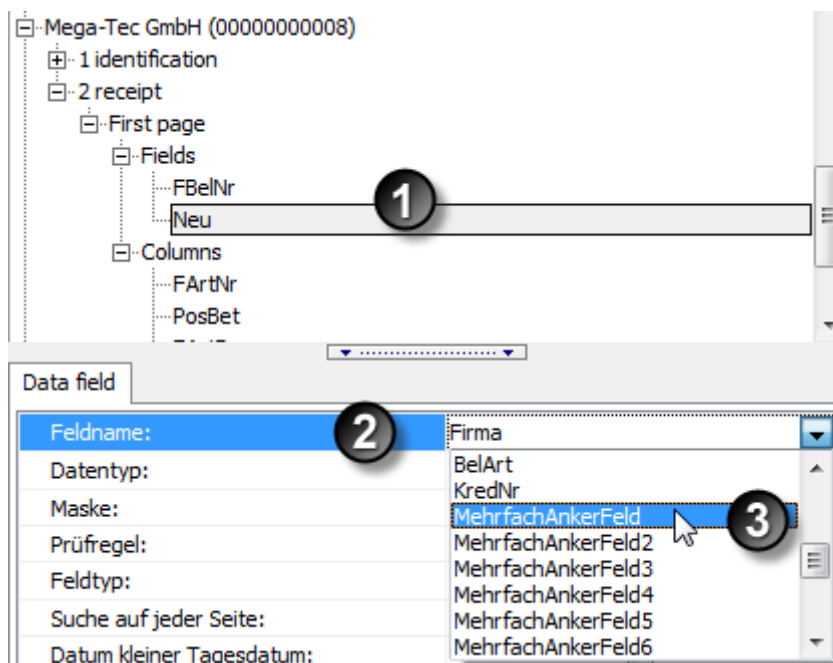


Figure 54: Multiple anchor field

Document reader
Create templates

Mautgebühren

Produkt	Anzahl der Pakete	Gesamtpreis	Steuerfreier Anteil	USt 19%
DPD classic (HF)	24	2,88 EUR	0,00 EUR	2,88 EUR
Summe Maut:	24	2,88 EUR	0,00 EUR	2,88 EUR

Legende:
MDC = nicht paketgebunden

Produkt	Gesamtpreis	Steuerfreier Anteil	USt 19%
Serv. Fax/Datatax	5,56 EUR	0,00 EUR	5,56 EUR

24 Positionen	Netto steuerpfl.	USt 19%	Netto steuerfrei	Brutto
Gesamtkosten	96,93 EUR	18,42 EUR	0,00 EUR	115,35 EUR
Rechnungs- / abschläge	10,77 EUR	2,05 EUR	0,00 EUR	12,82 EUR
Finanzierungskosten (Fiko)	5,54 EUR	1,05 EUR	0,00 EUR	6,59 EUR
Rechnungsbetrag o. Fiko	110,70 EUR	21,03 EUR	0,00 EUR	131,73 EUR
Rechnungsbetrag mit Fiko	116,24 EUR	22,09 EUR	0,00 EUR	138,33 EUR

Speditions-Rg. sind sofort und ohne Abzug fällig.

DPD erbringt Leistungen nur nach Maßgabe der Allgemeinen Geschäftsbedingungen der DPD Dynamic Parcel Distribution GmbH & Co. KG...
Geschäftsführer: Bernd Brödel, Greg Hallman, Jörg Lötcher, Ute Schür...
Kontakt: +49 10 75208999

Figure 55: Example of where the field can be located

In our case, the value of the overall cost is to be exported. Depending on the number of positions, the line on which the value is located can shift up or down.

The yellow marking specifies the area (column) (1) in which the anchor text (in our case, the overall cost) is found. Enter the anchor text (2) and % match (3).

The screenshot shows a document page on the left and a configuration window on the right. On the document page, a yellow box (1) highlights the 'Gesamtkosten' entry in the table. The configuration window has the following fields:

- Feldname: MehrfachAnkerFeld
- Ankertext: Gesamtkosten (2)
- Übereinstimmung zu: 70% (3)

Figure 56: Define area

Complete this first step by clicking **Save**, then click the button **New** and define the field (4) for the anchor.

Document reader
Create templates

The screenshot shows a document viewer on the left and a configuration sidebar on the right. In the document, a yellow box labeled '5' highlights a table with columns: 'Netto steuerpfl.', 'USt 19%', 'Netto steuerfrei', and 'Brutto'. The sidebar on the right shows a tree view where 'MehrfachAnkerFeld' is selected and highlighted with a red box and a circle labeled '4'. Below this, the configuration for 'Mehrfach-Anker-Feld' is visible, including fields for 'Feldname', 'Datentyp', 'Prüfregel', 'Maske', and 'Suche auf jeder Seite'. The 'Datenmanipulation' section is also expanded, showing options like 'Ersetzen', 'Leerzeichen entfernen', etc.

Figure 57: Define area

The yellow area (5) is approximately as large as the area for setting the anchor. The column signals the document reader that the value to be exported is within the column and **on the same level/line** as the anchor. The only thing that is left is the known definitions for the field name, data type, etc.

16. Special case: Multiple anchor fixed field

The screenshot shows the 'Datenfeld' configuration window. It has a list of field types on the right side. The selected field type is 'MehrfachAnkerFixFeld1'. Other visible field types include 'Firma', 'MehrfachAnkerFeld8', 'MehrfachAnkerFeld9', 'MehrfachAnkerFeld10', 'MehrfachAnkerFixFeld2', 'MehrfachAnkerFixFeld3', 'MehrfachAnkerArrayFeld', and 'MehrfachAnkerArrayFeld2'. The 'Datenmanipulation' section is also visible at the bottom.

Figure 58: Multiple anchor fixed field

The multiple anchor fixed field can be employed if the multiple anchor field cannot be used because the field to be exported is not on the same line as the anchor and is instead at a fixed location depending on the anchor. A column with an anchor is defined along with the value to be exported in the multiple anchor fixed field (within the definition *Fields*).

Document reader
Create templates

In our case, the value to be paid is exported in EURO (1), and the term Payment conditions (2) is the anchor. First the column is identified in which the anchor text is found (yellow marking). The term can be anywhere within this column (see the red marking). From there, the amount that is to be exported is selected (see the blue marking).

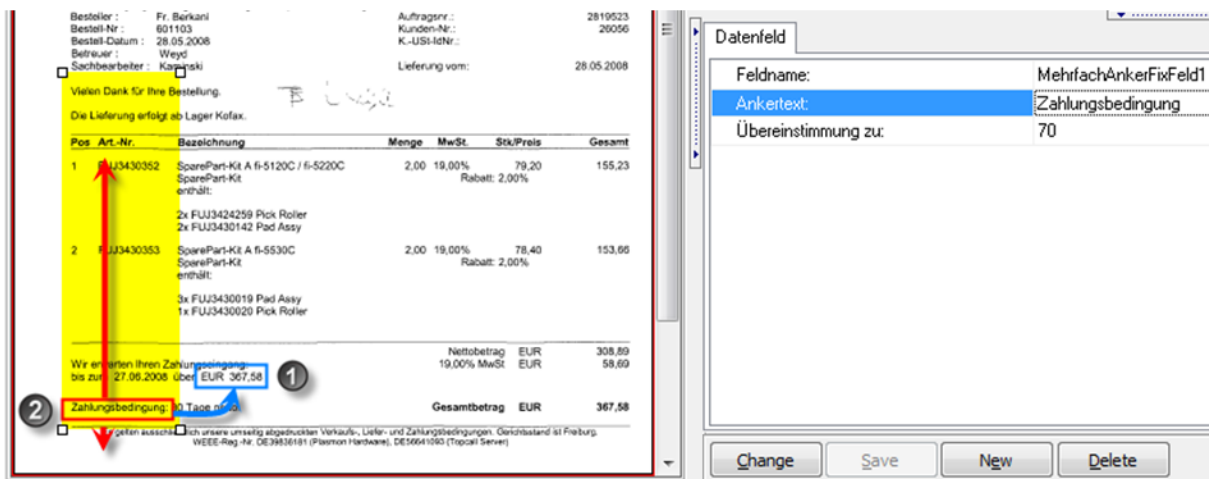
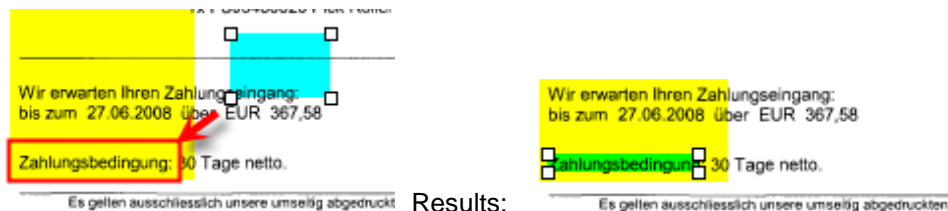


Figure 59: Definition of areas

In addition, the anchor text needs to be defined with turquoise area. Create this area closely around the term to prevent misrecognition.



Initial situation: Es gelten ausschliesslich unsere umseitig abgedruckten

Results: Es gelten ausschliesslich unsere umseitig abgedruckten

Then click the *New* button to define the associated fixed field. In contrast to the multiple anchor field, the fixed field is marked as a specific area and not as a column (make sure to place the box closely around the term or, if necessary, leave space for more characters). As with other fields, add the field name, data type and data manipulation, if applicable.

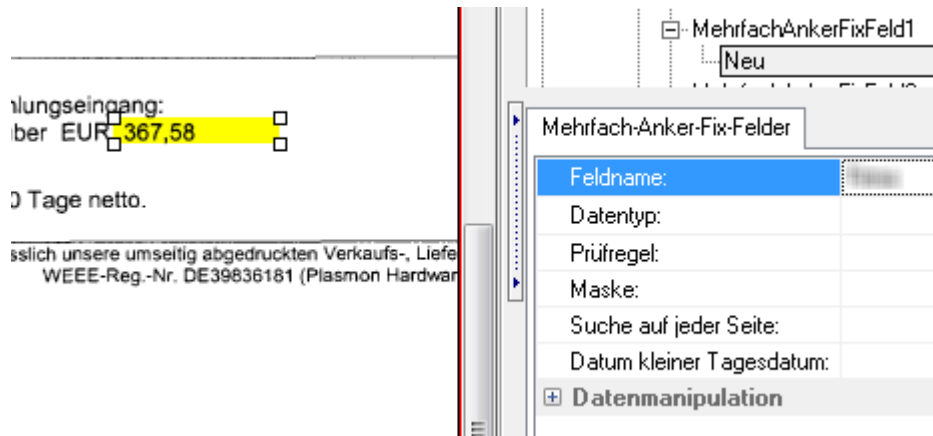


Figure 60: Definition of a fixed field

17. Special case: Dummy

Select dummy (within the definition *Columns*) as the field name when a database is not to be filled with the exported value, but the field is relevant for the evaluation of saved rules (scripts). A dummy is used, for example, when the following rule (script) is saved: Article items are invalid when the Amount field is empty (the amount is not a database field in our example and is therefore read out as a dummy).


18. Special case: Negation of values

For identification, certain rules can be saved for areas in the template that apply when the specified value is not in the selected location (in this area) (for example, it is a certain supplier, or the template is for a certain supplier). Select the area and save the term with a preceding \$ in the verification rule.

19. Import/export folder

To provide HABEL with the entire folder including the scanned stacks and recognized values (for example to troubleshoot problems), the folder can be exported.

A separate program (hphab405) is available that exports the data for the transfer in the HABEL directory folder Transfer/Export as a zip file dated with the current day. This folder is sent to HABEL.

	<p>The <i>Import</i> function is only provided after <u>contacting HABEL</u> or for test purposes in the test system. During importation, the data that existed at the time of exportation overwrite the current data.</p>
---	--

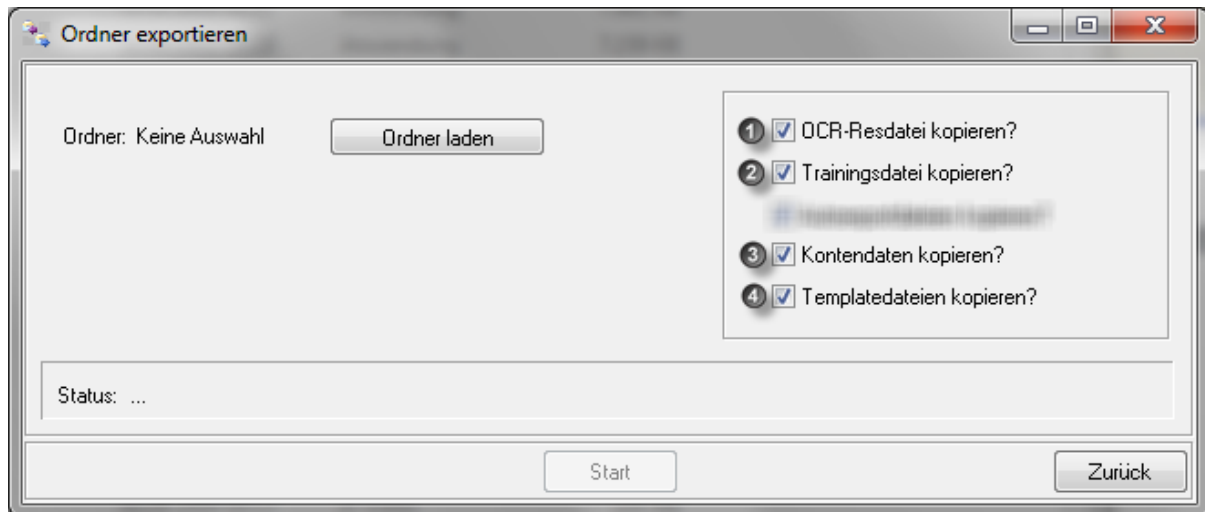


Figure 61: Exporting folders

Different parts of the document reader can be exported:

(1) Copy OCR RES file	The raw text/OCR result is exported in the original state
(2) Copy OCR training file	All of the identifiers and features that were trained for the document reader are exported.
(3) Copy content data	The master data in the document reader are exported.
(4) Copy template files	Saved templates are exported.

Initiate folder selection by clicking the button Load folder, and select the folder that is to be exported. Click Start to export the checked files to the indicated directory. You can then send the zipped file to HABEL. With this file, HABEL can reconstruct your current system including the data and help you troubleshoot.

**"Stop improving and
you cease to be good!"**

