

ALLES GEHABELT. ALLES GEREGLT.

Document reader

Software manual



Release 2018

Table of contents

1	Introduction.....	4
1.1	Preface.....	4
1.2	Overview.....	4
1.3	Symbolism in this manual.....	4
1.4	Information regarding this manual.....	4
2	Variants of the document reader.....	5
3	Applications of the document reader.....	6
3.1	Read document information and provide values.....	6
3.2	Read with interpretation and counter-checking of data.....	7
3.3	Read with subsequent editing in the post basket.....	8
3.4	Automatic archiving of outgoing documents.....	10
4	Document capture/recognition.....	11
4.1	General.....	11
4.2	Scan documents.....	11
4.3	Recognition process.....	12
4.4	Recognition of the creditor/debtor.....	13
4.5	Read the document information with free-form recognition.....	17
4.6	Check whether a suitable template exists.....	18
4.7	“Create processes”.....	19
4.8	Correction of recognized data.....	22
4.9	Further commands in the “Create processes” program.....	23

1 Introduction

1.1 Preface

The HABEL document management system is a database-supported administration system for electronic documents.

The purpose of this software manual is to assist you in performing responsible work with HABEL.

If you have any questions about the system, our support staff will be glad to help you.

1.2 Overview

This user manual is aimed at users who work with the document reader. It explains the specific functions and settings.

1.3 Symbolism in this manual

General symbols

➔ This list marks the description of actions for you to perform.

- This item marks bulleted lists.
- This line marks lists on the second level.

Cross references

Cross references are displayed as follows: ↪ *“General symbols” on page 4*

1.4 Information regarding this manual

The figures may differ in certain details from your HABEL document management system, as functions may be included that have not been purchased and/or activated for your system. These can generally be easily added to your system. Please speak to your customer support agent about this.

Please send any suggestions, requests or criticisms to **handbuch@habel.de**. Thank you.

The captions and specifications in this manual are only used as examples and therefore can't be applied 1:1 to every situation. For that reason, you should consider your situation individually.

2 Variants of the document reader

There are three variants of the document reader which differ in terms of their functions. The right variant for you is determined on the basis of your requirements. The applications of the document reader (in our case, the 3rd variant) are explained in the information that follows.

The table below shows the differences between the variants:

Function	DOCUMENT READER BL50	DOCUMENT READER BL100	DOCUMENT READER
OCR	Yes	Yes	Yes
ICR	Yes	Yes	Yes
Max. volume of templates.	Max. 50	Max. 100	Unlimited
Reading of data with templates.	Yes	Yes	Yes
Barcodes must be affixed for separation of procedures.	Yes	Yes	No, but possible
Data read	Header and footer data max. 10	Header and footer data max. 20	Unlimited
Item recognition	No	No	Yes
Automatic invoice checking.	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
■ Checking of recognized data against external sources and other interpretation of data.	No	No	Yes

3 Applications of the document reader

3.1 Read document information and provide values

The document reader reads document information and provides values for further processing. Below is a graphic representation based on the example of a purchase invoice:

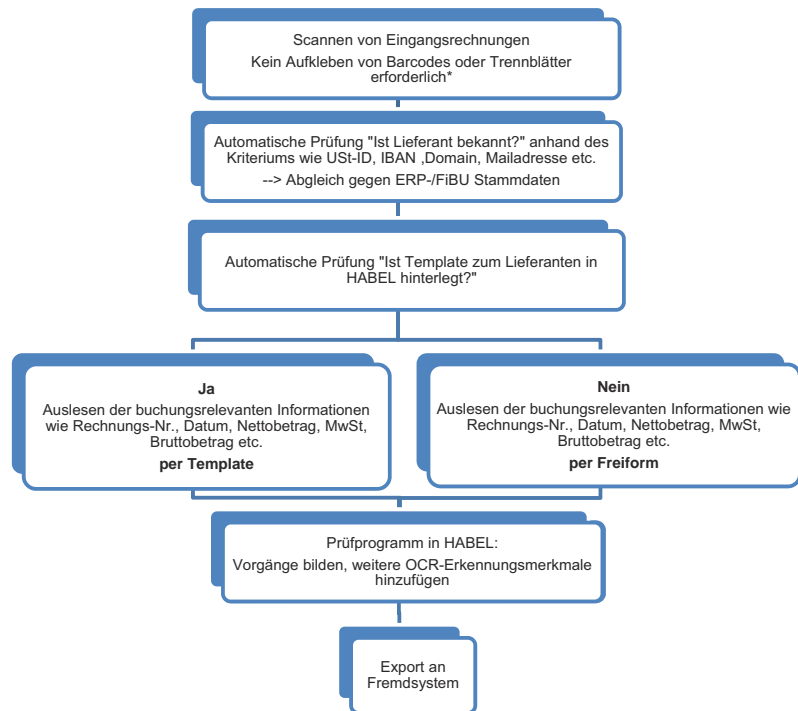


Fig. 1: Read the document information for a purchase invoice

* Not in conjunction with document reader BL50/document reader BL 100.

3.2 Read with interpretation and counter-checking of data

Data such as currency abbreviations etc. can also be interpreted and added. Data can be counter-checked against ERP/financial accounting data for automation of procedures such as posting clearance. Below is a graphic representation based on the example of a purchase invoice:

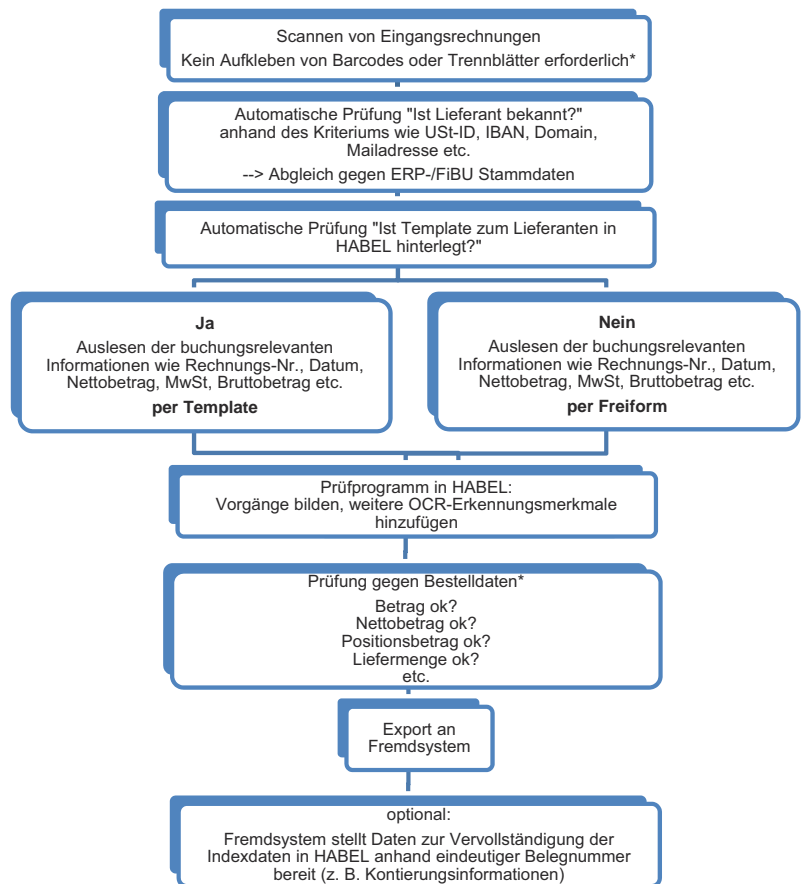


Fig. 2: Read and counter-check the document information for a purchase invoice

* Not in conjunction with document reader BL50/document reader BL 100.

3.3 Read with subsequent editing in the post basket

If discrepancies are discovered when counter-checking, the data can be edited manually in the post basket. The post basket can also be used for basic editing (workflow). Below is a graphic representation based on the example of a purchase invoice:

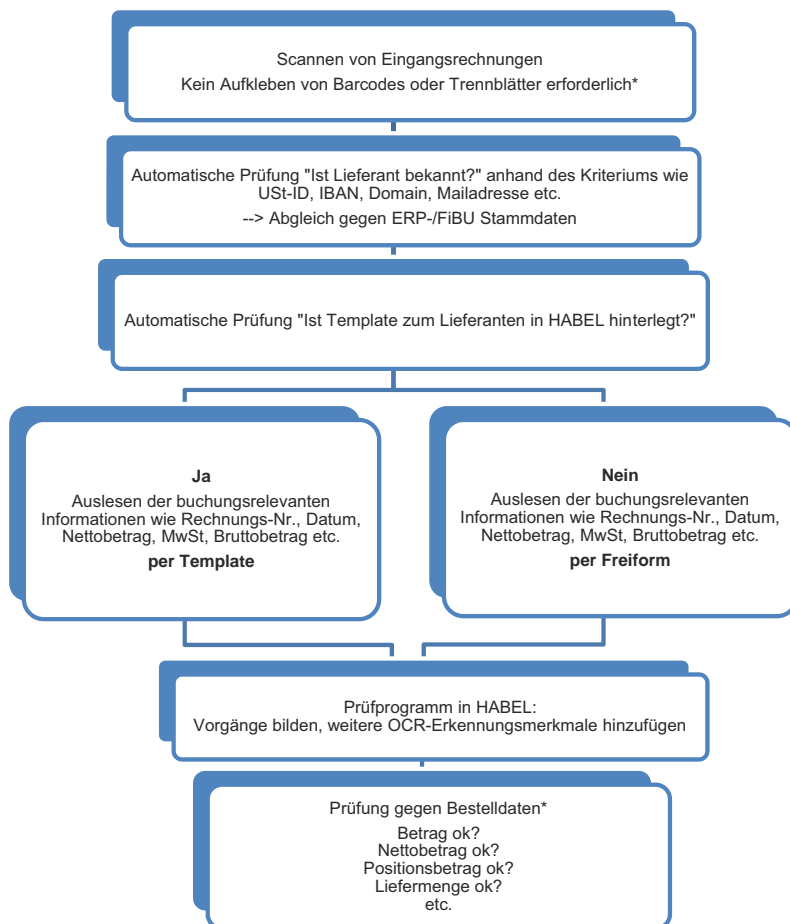


Fig. 3: Read the document information with subsequent workflow

* Not in conjunction with document reader BL50/document reader BL 100.

Continued:

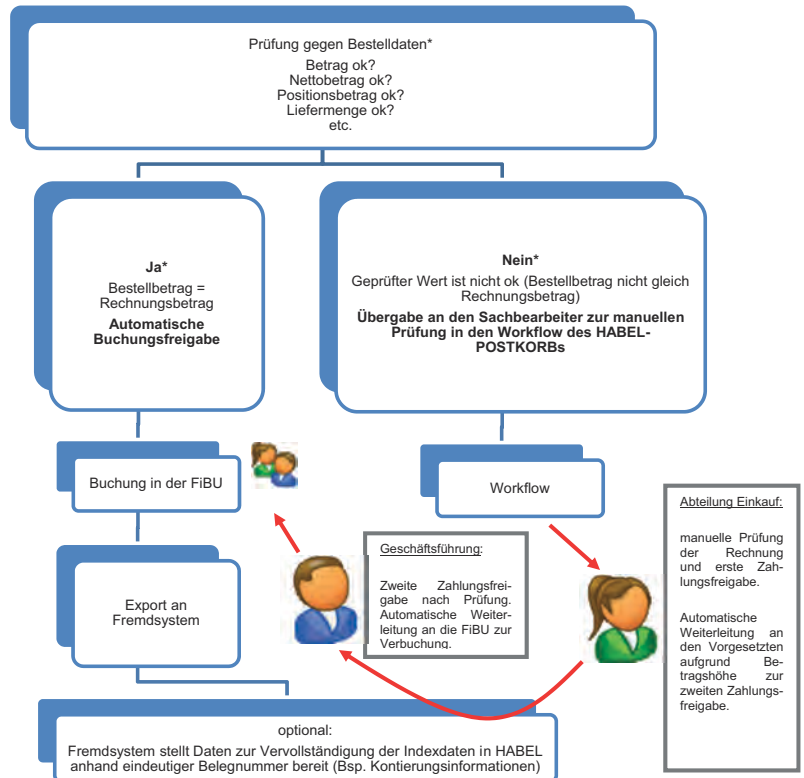


Fig. 4: Workflow of a purchase invoice

* Not in conjunction with document reader BL50/document reader BL 100.

3.4 Automatic archiving of outgoing documents

The document reader can also be used for the archiving of outgoing documents supplied without index data. The chart below shows the process for this:

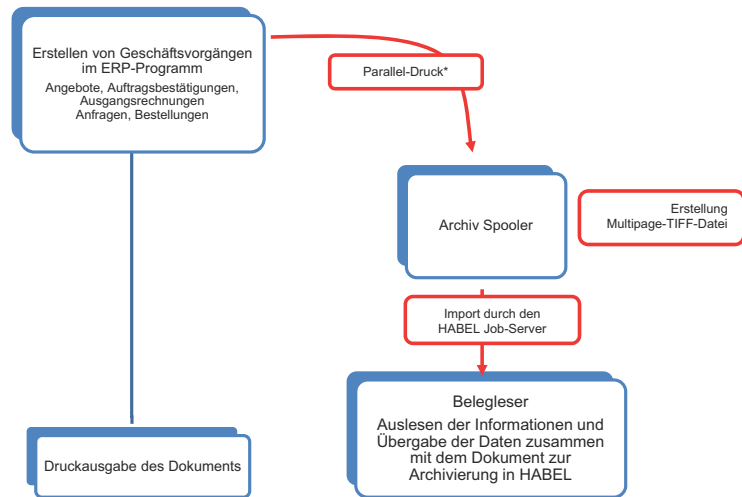


Fig. 5: Automatic archiving of outgoing documents

* Advantage: With this printing process, all the operations are performed digitally, so the document quality remains consistently good for reading by the document reader, leading to almost 100% recognition.

4 Document capture/recognition

4.1 General

With the document reader, information can be read directly from documents during scanning and used for processing.

The reader has many applications such as scanning and editing of **customer orders** or **purchase delivery notes**. This module is frequently used for editing of **purchase invoices**. The descriptions and associated terms in this manual are based on editing of purchase invoices. However, they can also be applied to other areas.

Various separate programs are available for using the document reader. Users will mainly use the familiar "KOFAX" ("HpKfx100") scanning program and the program for **process creation** and **document reader training** ("HpHab401"). Administrators can access other programs, which will be described later. The programs can be found in the HABEL server directory under **\GEDOSOD\32-Bit- directory**.

An **OCR license** ("ABBY") is required to read the raw text of the documents. A license is required in all places where the document reader is used.

4.2 Scan documents

The documents are captured in a stack with the "KOFAX" scanning program. With the exception of the document readers 50 and 100, the procedures do not need to be separated by barcodes or separator sheets.

The documents which belong to a procedure are defined on the basis of a criterion. In most cases, the document number is used for this.

Mandant:	01	Aufzug HABEL GmbH & Co. KG
Belegart:	999	OCR Eingangsbelege
Ordnerdatum:		

Fig. 6: Scan program

Select the document type which has been set up for processing with the document reader. This also defines which fields must be input for scanning. This is usually just the **folder date**, i.e. the date on which the current stack (folder) is scanned. If scanning is performed every day, then the current date must be entered in each case. If scanning is performed for multiple days in the past, then this date can be used for differentiation. The documents are then scanned in a stack.

4.3 Recognition process

Once the scanning process is complete and the data has been transferred, recognition of raw text can take place. This is usually initiated immediately after scanning (exit, transfer). Optionally, the members of staff in charge of the process can be informed when recognition is complete.

The “HpHab400” program can also be started manually.

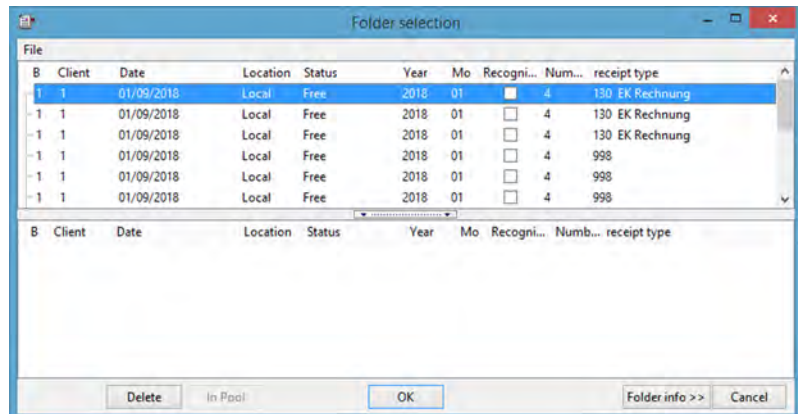


Fig. 7: Folder selection

The scanned stack (folder) with Date, State, Year, Recognized and the Number, Document type etc. is displayed in the area at the top.

If incorrect stacks have been scanned, you can delete them with the [Delete] button. They will then not be found in the archive either. After selecting the stack, start recognition by clicking [OK] and then [Recognize all].

For recognition of document information, the following steps are performed:

- “Recognition of the creditor” using free-form recognition
- “Reading of the document information” using free-form recognition.
- “Checking” whether a “template” exists for the creditor.
- “Reading of the information” via the template (if necessary, overwriting of the values detected with free-form recognition).

4.4 Recognition of the creditor/debtor

The creditor/debtor can be recognized in various ways. Features stored for each customer system are used by the document reader to identify the creditor/debtor. These features are called “*identifiers*”.

Identifiers include:

- E-mail address
- URL
- Customer number
- Tax number
- IBAN
- Account number

Values are stored for this identifier for each creditor/debtor, which enable the creditor/debtor to be identified. The document reader reads the raw text for the document and compares it with the stored identifier.

If the “*identifier is recognized*”, the system searches for the “*corresponding value*” in the immediate vicinity of the term (clock-wise: right, down, left, up). This is how “*allocation*” to the creditor/debtor takes place. The stored templates are also checked to determine whether allocation to the creditor/debtor can be performed based on the identification features stored there.

Recognition of identifier variants — Training the document reader

A range of terms that can be used as identifiers are pre-stored in the system.

However, there are often different spellings or words which need to be used as identifiers. These variations can be stored in the system. The document reader is trained to recognize them. This training is beneficial, because it means that these values will be recognized as correct in future which gradually reduces the amount of manual intervention.



NOTICE!

This becomes problematic primarily in the case of terms which are not unique, e.g. date — document date. For some creditors/debtors, the date on the document is always the document date, but for others it might be different date information such as the delivery date.

An extension of the identifier might be useful for the term Account, for example, as this is used in different ways:

- Account number
- Account no.
- Acc. no.
- Account

If you have defined that the account number is a recognition feature for creditors/debtors and therefore an identifier, you should also store the various **term variants**.

Add/delete identifier

You can add further identifiers in the context menu of the document display. In our example, we are adding the term variant “*Account:*” to the identifier “*AccNo*” which has already been stored. To do this, right-click on the term variant that you want to add.

The context menu opens. Navigate to the menu item “*Add identifier* → *Creditor* → *Acc. no.*”

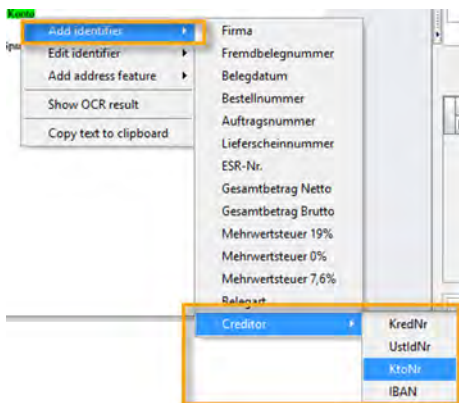


Fig. 8: Add identifier (context menu)

A window opens for entering the “*identifier name*”. The field is pre-allocated with the name of the term variant to be added. You can change this manually if you wish.

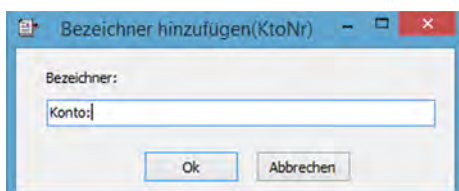


Fig. 9: Identifier name

If you want further values recognized from the document to be transferred, you can do this by clicking on the term while holding down the [SHIFT] key. Select “*Edit identifier*” in the context menu to view and, if necessary, delete the stored term variants for the identifier in question.

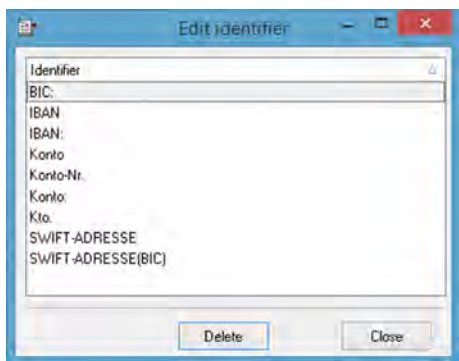
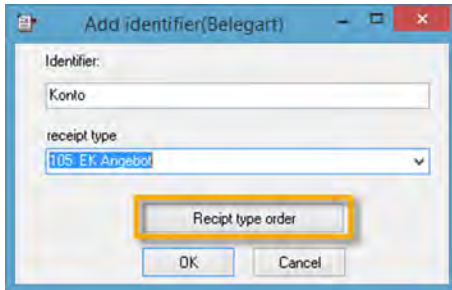


Fig. 10: Edit identifier

Define identifier for document type

If document type recognition is also required, you need to train the document reader in this case, too. The documents are scanned with a dummy document type and recognized as the corresponding document type based on particular identifiers.



Firstly, the identifier is determined for the document type and the corresponding document type defined.

Use the document type order to define which document types can be linked to this identifier. For example, the term “Credit note” appears on a credit note, and the word “Invoice” is often on it, too. If the term “Credit note” is closer to the top, the document is a credit note. This is defined in the document type order. To define the order, move the document types as required.

Fig. 11: Add identifier (document type)

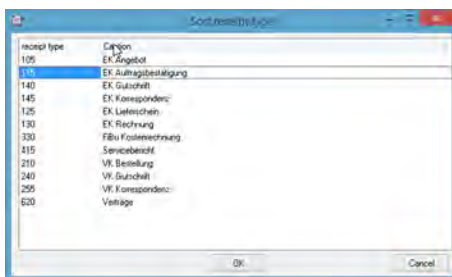


Fig. 12: Sort document type

Add address features

To recognize the creditor/debtor, the identifier is read from the raw text and the corresponding value in close proximity to it.

Features for this are stored for the creditors/debtors. If the creditor/debtor is recognized, the fields to be filled in are completed using data from the address master data. If the creditor/debtor is not recognized (field not filled in), the master data in the ERP program must be added. The data from the ERP program is transferred to the HABEL address master data and is available for further processing in HABEL. The ERP program does not have access to the data stored in the document reader.

Right-click on the document to open the context menu. This gives you the option to add address features. It is ideal if you right-click on a recognized value that is significant for this creditor/debtor. In this case, the account number (1).

The stored terms need to be individually defined for each customer system and may therefore be different from your system. After selecting an address feature, you will be taken to the capture mask.

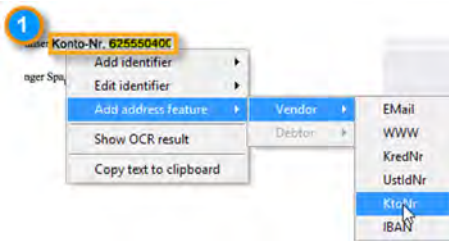


Fig. 13: Address features context menu

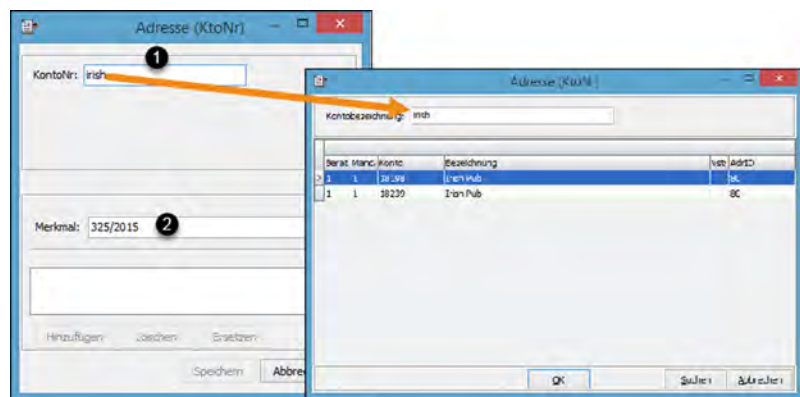


Fig. 14: Address feature with master data access

The name of the creditor/debtor is entered in the “Account No” field (1). The recognized feature has already been transferred to the “Attribute” field (2). Use the [Tab] key for master data access. This lets you access the data stored in the address master data.

Click the [Search] button (3) to check whether master data already exists. If it does, it can be transferred. Recognition of the creditor/debtor is broadened to include the newly stored feature, provided this is added as an attribute.

4.5 Read the document information with free-form recognition

Free-form reading principle

The HABEL document reader has various stored words which may appear on the archived documents and which indicate certain content.

If one of these terms is found on a page, the document reader will search the area surrounding this term for eligible values. Firstly, it searches to the right of the term. If it does not find any eligible values here, it continues to search clockwise around the term. These terms are called “*identifiers*” in HABEL. Whether a value is accepted as **eligible** depends on the type and/or structure of the recognized value. The system is set up with logic to recognize whether a value is eligible or not.

For example, the term “*Invoice no.*” is stored as an identifier for the invoice number. If this term is found on the document, the document reader will then search for a number or a combination of digits next to it, which is accepted as the invoice number.

If it finds a value of this kind, this will then be transferred directly to the corresponding database field. If it does not find a value or the value it finds is not accepted (because it is a date, for example), it will continue to search the area surrounding the term.

The stored terms are used for all documents and all suppliers. This method allows very large numbers of values to be recognized with relatively little effort.



Fig. 15: Search schema

Free-form training

When the document reader is launched, the most common terms are normally pre-stored as identifiers. However, these terms only represent a basic selection and can be added to as required. By adding further identifiers, you can train your document reader and further improve the reading results. To train it to recognize new terms, right-click on the document to open the context menu and select “*Add identifier*”. Then select the database field to be used, for which the term will be an identifier.

If the cursor is positioned on recognized text before you right-click, this text will be transferred directly to the dialog box which then opens.



Fig. 16: Free-form training

The text can then be changed, added to or entered completely manually in this dialog box. By confirming with [OK], you train the document reader to recognize this term as the identifier for all documents for the selected database field. If you have trained the reader with incorrect values or if you want to check what values it has been trained to recognize, you can call up all terms stored for a particular database field. To do this, right-click on the document and select “*Edit identifier*” in the context menu. You can easily delete incorrect values by clicking “*Delete*”.

Limitations of free-form recognition Many values can be transferred to the correct database field using this method. However, there are certain values that cannot be read correctly in this way. A few examples of instances which are beyond the capacity of free-form recognition are listed below:

No identifier present	With date information, it is frequently the case that the date is printed without an identifier. If there is no identifier, this value cannot be read.
Identifiers are too general.	The identifiers need to be sufficiently precise so that they can be recognized as identifiers. Identifiers that are not sufficiently precise often lead to incorrect results. For example, the word "Date" without any additional term is very imprecise. It is not clear whether it is an invoice date, a dispatch or delivery date, or a sales order or purchase order date. If the identifier is trained to recognize this term, while this may yield correct results for one document, it will also lead to incorrect results for lots of others. It is up to you to decide whether it is useful to train it to recognize this term, but you should certainly make this decision prudently!
Distance too great	When it recognizes an identifier, the system searches for an eligible value to the right of the identifier first. The distance between the identifier and the value must not be too great. Particularly in the case of amounts, the identifier is often left-aligned and the value right-aligned. These values cannot be read with free-form recognition.
Incorrectly recognized values	With free-form recognition, the recognized values cannot be tampered with. If a value is recognized incorrectly by OCR recognition, this value cannot be automatically corrected. This is only possible in templates.

4.6 Check whether a suitable template exists

If a creditor has been recognized with free-form recognition, the templates of this creditor are checked in succession to determine whether one of them is suitable for the current document.

If a creditor has not been recognized yet, all the templates will be checked. As soon as a suitable template is found, the creditor is assigned via the template.

The values recognized with free-form recognition are overwritten with the content from the templates. If a value is recognized with free-form recognition and also read via the template for the same database field, therefore, the content of the template is always used. If, on the other hand, no content from the template is used, because a field has not been created in the template for this database field, for example, the values from the free-form recognition are retained.

4.7 “Create processes”



Fig. 17: Document reader — “Create processes”

The documents are displayed for further editing in the “Create processes” program (“HpHab401”).

You can decide which stack you want to edit in the folder selection. Folders which are being edited or stacks still to be recognized are listed in the bottom section of the folder selection.

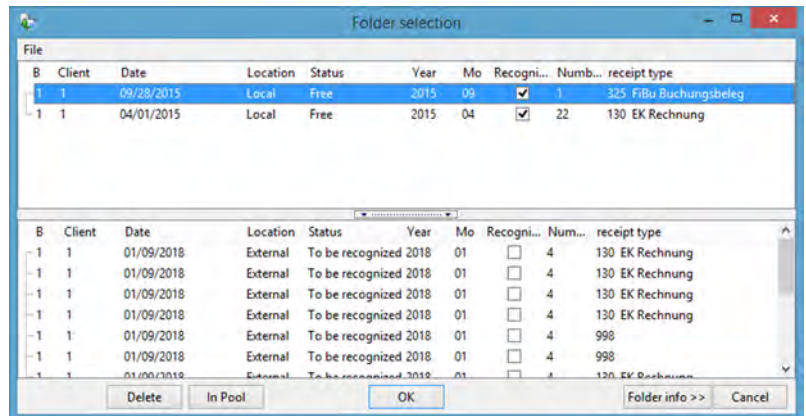


Fig. 18: Folder selection to “create processes”

You can also delete whole stacks here with the [Delete] button if they have been scanned by mistake.



NOTICE!

If a stack is deleted, the documents cannot be found in the archive, although data was transferred after scanning. The documents are only in the archive once editing with the document reader is complete.



NOTICE!

Even if the raw text has already been recognized, the workstation where the documents are processed with the “Create processes” program must have access to the OCR licenses (local ABBYY installation with access to the server). However, there is no consumption of sheet volume at this workstation.



Fig. 19: “Create processes”

After selecting the folder, the documents and recognized data are displayed in the program. You will see a document view on the left (1) and a hitlist of the recognized procedures at the top right (2). In the area to the right of center, all the data for the entry is displayed. In this area, you can also make changes such as enter values, access master data or delete the whole record (3). You can do the same for items. The items are displayed in table form (4) and can be edited individually (change, add, delete) by selecting them (5). The editing fields for the whole entry and the items can be shown or hidden (6).

If the rows in the hitlist are the same color with the same shading, the document reader has recognized this as a procedure. Each new procedure is shaded differently for a clearer overview. Procedures marked in green have been recognized as a procedure after recognition by the document reader based on the defined criterion (e.g. document number). If entries are marked in red, this indicates that it was not possible to determine whether the page displayed is the first page of a new procedure or whether it belongs to the procedure before or after.

If an entry only has one page and is marked in red, the criterion (in this case, document number) needs to be assigned to this entry. To do this, highlight the entry and press the *[Combine]* (9) button. You will be asked for the criterion.

Enter the criterion, and press the *[Create processes]* (8) button. The entry will then appear as correct (marked in green).

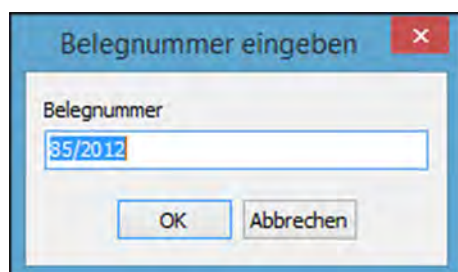


Fig. 20: Enter document number

In the case of multi-page entries that have not been recognized as a procedure, highlight all the entries that belong together and press *[Combine]* (9). You will be asked for the correct document number, which can then be entered.

In this case, too, press the *[Create processes]* button (8) at the end so that the result will be shown in green.

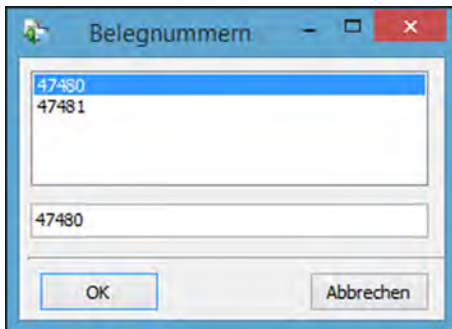


Fig. 21: Assign document numbers

If multiple document numbers have already been recognized for the entries that form a procedure, a selection window appears after pressing the *[Combine]* button (9) where you can assign the correct document number.

The selected document number is considered the correct document number and is assigned to all entries in the procedure after you press *[OK]*. The procedure will be shown marked in green after you press the *[Create processes]* button (8).



NOTICE!

Even if procedures have been combined and the *[Create processes]* button (8) pressed, the entry may still be marked in red. This is because there are other entries before or after it that the system has not yet been able to assign correctly. You can, of course, also only press the *[Create processes]* button (8) once all entries have been combined.

If process creation is complete, the recognized data can be exported for further processing by clicking *[Export]* (7). Entries marked in red will not be exported.

Further processing could, for example, comprise displaying procedures in the post basket that are not correct according to stored checking rules and which need to be manually checked, e.g. certain fields need to be filled in that were not filled in during document recognition. Further processing could also comprise transfer of the recognized data to the Accounts department to post the procedures. Alternatively, the data could be fed into an external program's workflow.

4.8 Correction of recognized data

Recognized data can also be corrected manually and missing data added in the “Create processes” program (“HpHab401”).

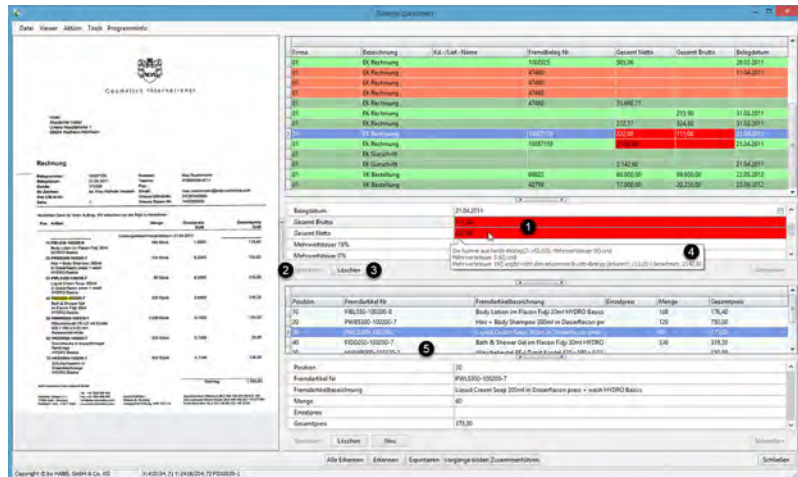


Fig. 22: Check — Change field values

The header data is shown in the top section and the item data in the bottom section (5). Click in the fields (1) to make changes directly. Use the [Save] button (2) to transfer the changed data. Which fields can be changed is defined. The [Delete] button deletes the whole record (3). Incorrect or unrecognized values are shown marked in red. Notification texts inform you of the reason for this (4). In our example, the plausibility check determined that net amount + VAT ≠ gross amount and therefore one (or more) value(s) has/have been recognized incorrectly. The same is possible for the item data in the lower section on the right of the mask (5).

Field values marked in the hitlist are marked in the document view at the same time so that you can quickly compare where the value was recognized and whether it is correct.

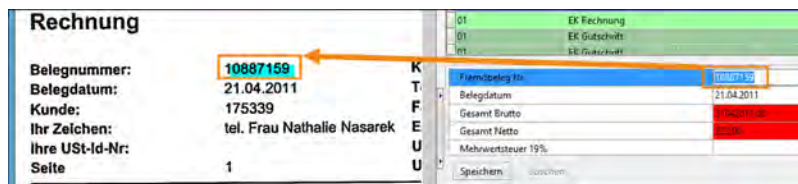


Fig. 23: Display of marked values

4.9 Further commands in the “Create processes” program

Context menu of the document view

- Recognize
- Copy text to clipboard
- Show OCR result

The command “Recognize” (1) can be used to start document recognition for this document again, for example, if new features have been added to recognize the creditor/debtor. The data is added when recognition is performed again.

The command “Copy text to clipboard” (2) makes the recognized value available for pasting in a location of your choice.

The command “Show OCR result” (3) displays the recognized values. They are grouped in Free-form result as well as Template (by page) and Header/item data.



NOTICE!

To display the results, start recognition again with the [Recognize] button after invoking the command.

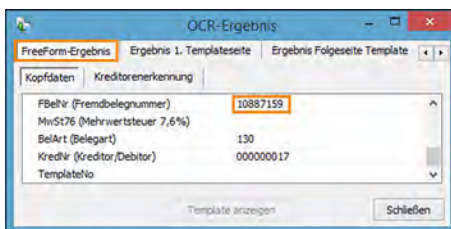


Fig. 24: OCR result

You will see an overview showing the origin of the data. The first tab shows all the values that were recognized with free-form recognition. These values were recognized, because a term that the document reader was trained to recognize was found in the document, this value was near the term and it corresponded to the stored rules.

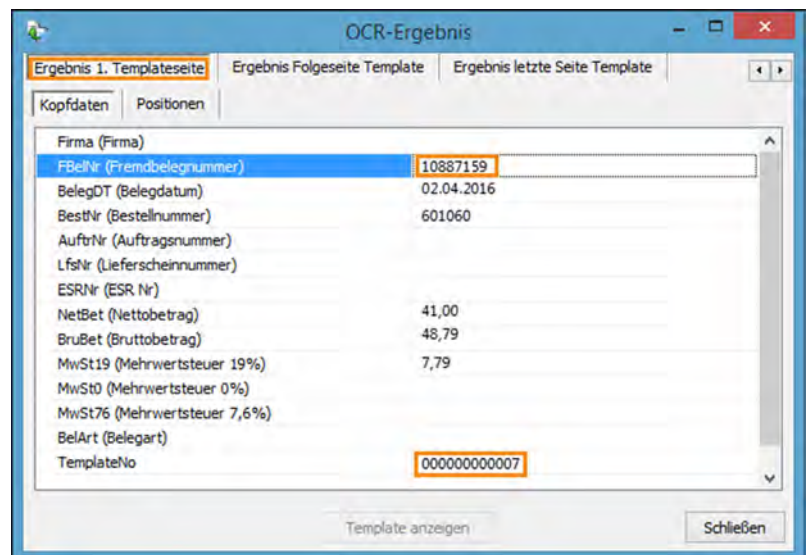


Fig. 25: Result 1 template page

If a template was created for the supplier or document in question, the values read via the template will be shown on the second tab (“Template page 1”). The values from the template overwrite the values detected with free-form recognition.

In this example, free-form recognition incorrectly recognized a 1 instead of an I in the external document number. By creating a template, incorrect recognition such as this can be automatically corrected.

If a template number is entered in the “*TemplateNo*” field at the bottom of this display, then the system recognized in all cases that a template exists for this document.

Even if no values can be read via the template, the template number used will always be displayed.

Press [*Show template*] to open the template used.

Context menu in the hitlist of recognized values

The commands “*Export*”, “*Recognize*”, “*Recognize all*”, “*Create processes*” and “*Combine*” are also available as buttons in the footer of the mask. Their use has already been described. The data that can be called up with the command “*Show OCR result*” has also been explained in the previous section.

The context menu also includes the following commands:

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

The command “*Load folder*” (1) is used to call up the folder selection and to call up further stacks for editing in the “*Create processes*” program.

If scripts are stored such as plausibility checks of values or separation logic, these can be started with the “*Run script*” command (2). If the command is not used and the command “*Create processes*” is executed at the end, the scripts will be run at this time.

The command “*Change field value*” (3) lets you change the field you have right-clicked on. An input mask appears in which the required value can be entered directly.

The command “*Delete document*” (4) deletes the selected document.



NOTICE!

After this, the document is not present in the archive.

Menu bar

The menu bar offers various commands that you are already familiar with. Other commands are explained briefly below:

<i>"Print TIFF file"</i>	For printing the document
<i>"Save document as"</i>	Can be used for creating templates.
<i>"Viewer"</i>	Customization of the document view and information about the TIFF file.
<i>"ASCII chart"</i>	Display of the usable characters in ASCII format.
<i>"Debug window"</i>	Further details regarding script creation.



DEUTSCHLAND

HABEL GmbH & Co. KG
Untere Hauptstraße 1-5
D-78604 Rietheim-Weilheim
Fon +49 7461 9353-0
Fax +49 7461 9353-99
www.habel.de | info@habel.de

Niederlassung Leipzig
Messe-Allee 2
D-04356 Leipzig
Fon +49 341 678-27322
Fax +49 341 678-28322
www.habel.de | info@habel.de

SCHWEIZ

HABEL Dokumentenmanagement GmbH
Rheinstrasse 36
CH-8212 Neuhausen am Rheinflall
Fon +41 52 674-8151
Fax +41 52 674-8150
www.habel.ch | info@habel.ch