

docWizz

User's Manual



Contents

- 1 Introduction..... 4**
- 2 Workflow overview 6**
- 3 Description of docWizz user interface 8**
 - 3.1 Bars 11
 - 3.1.1 Workflow bar..... 11
 - 3.1.2 Page bar 14
 - 3.1.3 Status bar 20
 - 3.1.4 Menu bar..... 20
 - 3.1.4.1 Document menu 21
 - 3.1.4.2 View menu 22
 - 3.1.4.3 Page menu 23
 - 3.1.4.4 Page image menu 24
 - 3.1.4.5 Zone menu..... 26
 - 3.1.4.6 Detail menu..... 26
 - 3.1.4.7 Configuration menu 28
 - 3.1.4.8 Help menu 29
 - 3.2 Tools in different views 31
 - 3.2.1 Image view..... 31
 - 3.2.2 List view 36
 - 3.2.3 Text view..... 37
 - 3.2.4 Tree view 40
 - 3.2.5 Metadata view..... 41
 - 3.2.6 Properties view 42
 - 3.2.7 Clip view 43
 - 3.2.8 Custom view 44
 - 3.3 Rescan..... 45
 - 3.4 Explanation of workflow steps 49
 - 3.4.1 Process documents 49
 - 3.4.1.1 Document pool..... 52
 - 3.4.1.2 Go to 62
 - 3.4.1.3 Merge (Stitching) 63
 - 3.4.1.4 Knife - Polygon 65
 - 3.4.2 Import..... 68
 - 3.4.2.1 How to use the Setup import task 68
 - 3.4.2.2 How to use the Review import task 73
 - 3.4.3 Cropping 75
 - 3.4.3.1 How to use the Prepare cropping task (basic) 77
 - 3.4.3.2 How to use the Prepare cropping task (advanced) 78

3.4.3.3	How to use the Review cropping task	91
3.4.4	Zoning.....	91
3.4.4.1	How to use the Review zoning task	91
3.4.4.2	How to use the Review page sequence task	105
3.4.5	Structure	111
3.4.5.1	How to use the Review issues task.....	111
3.4.5.2	How to use the Review structure and text task	116
3.4.6	Output	157
3.4.6.1	How to use the Review output task	157
3.4.7	Rejects.....	157
4	docWizz Control Center.....	161
4.1	Configuration tool	161
4.2	Import document.....	163
4.3	Services status	164
4.4	Pool management	168
4.5	Storage capacity.....	175
4.6	Environmental control.....	176
4.7	Custom control	179
5	Remote QA (Quality assurance)	184
6	Backup, Autosave, Update.....	185

1 Introduction

Last updated: 17.03.2022



Congratulations!

We are happy to welcome you to the docWizz family. Thank you for purchasing docWizz, a system that enables you to easily digitize and convert valuable materials. This manual intends to explain docWizz in a simple manner so that you can get started quickly and see results.

As an user of docWizz you are usually confronted with a system that is ready to use. From time to time we will point out, that access to special program elements and functions depends on settings which can only be handled by the system administrator. Most of these technical elements and functions can be found in the Reference book.

docWizz is made to fit the specific needs of its users, because of this the following descriptions and diagrams may or may not completely match your docWizz configuration.

This manual has been undertaken a general review by the CCS team. If you, however, should find any inconsistencies, if you require further explanations, or you find that key questions are inadequately dealt with, we would be very grateful to hear your suggestions. Your suggestions are important to us for the improvement of our manuals.

Support and training sheets

More support and training sheets can be found in the CCS information center:

<https://content-conversion.com/>

Often used expressions

Monograph

This term is used to describe just about anything that is not a serial publication or newspaper. They are usually a paper, book, or other work about a single subject, and often written by a single author.

Newspaper

It is a type of publication which usually contains news, other informative articles, as well as advertising. These are typically published daily or weekly.

Serial

Refers to materials issued under the same title in a succession of parts usually numbered or dated, and appearing at regular or irregular intervals. The most common example would be magazines.

What are Rejects?

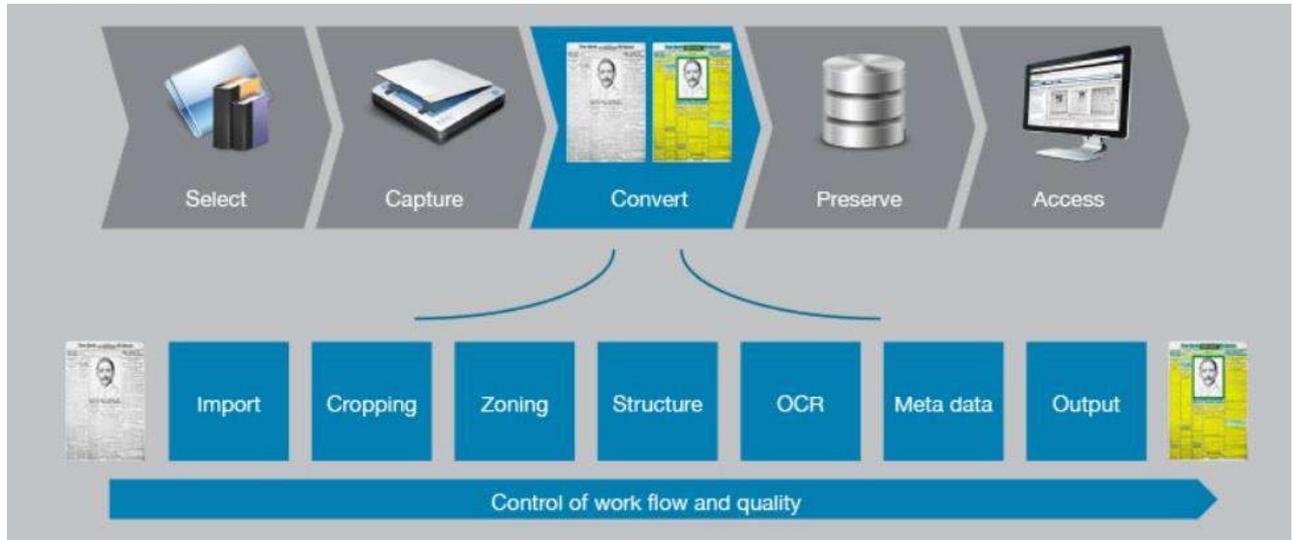
You may see the term "rejects" mentioned often in docWizz, and docWizz documentation, so let's talk about what they are. First of all, rejects are very specific to individual clients and even more specific to those client's project configurations. Think of rejects as a set of automated quality assurance rules. These rules will check certain parameters such as image resolution, OCR (Optical Character Recognition) confidence, missing pages etc. Each step in docWizz has its own set of rules, and if something does not conform to these rules a reject is created and will stop a document from processing any further until the operator takes appropriate action.

2 Workflow overview

Below is an illustration of a typical digitization workflow, each one of these steps could be discussed at great length but we want to focus on is the conversion process, which is where scanned images are ingested into docWizz.

It should be noted that the order of these steps can change on the type of document being processed.

dW environment



The digitization and conversion workflow with docWizz

Import	Here, images need to be selected for import. The appropriate project configuration is selected in this step but certain settings can be made on-the-fly as well such as OCR language and different analysis options.
Cropping	This step is used to crop images, clean borders, and can also be used to split double pages.
Zoning	In order for best searchability, accuracy and efficiency, the images are zoned according to their content (ex. Author, text block, Headline, Illustration, Page number, etc.) This is first done in an automatic analysis which includes both the physical dimensions of the zones and the type of zone.
Structure	Different kinds of structure are applied to different kinds of documents, for example books are divided up into Front, Main and Back sections. Various other structural elements are also arranged here, such as chapters or articles if it is a newspaper, as well as all of the zones identified in the previous step.

OCR	Optical Character Recognition This refers to the electronic conversion of images with the text into machine-encoded text. This is what will give your documents the power of searchability. There are a variety of different OCR engines that docWizz can use to accomplish this task.
Metadata	The final step is creating output according to the user's specifications, there are number of different formats that can be chosen and specified using the Configuration tool.

3 Description of docWizz user interface

After you have started docWizz using the icon on your desktop or from Start - Programs - docWizz, the system opens the standard user interface.

This interface is flexible. Some functions are accessible all the time without any changes; some change appearance and content depending on the working mode of the system; some are accessible only in special program situations.

The Welcome screen is displayed every time docWizz is opened, a document is closed, or if the Document Pool is closed. It will never be displayed while a document is open.

It can be closed using the close button or by pressing "Esc" key. If the Welcome screen is not displayed, press "Alt" key to display the menu, go to "View" submenu and make sure that "Show welcome screen" entry is checked.

Do not show again

Do not show again

If the "Do not show again" checkbox is checked, the Welcome screen will no longer be displayed. It can be activated by using the entry from the menu.

Create new document



Switches the task to "Setup import" in order to create a new document. If the current task is already at "Setup import", this button will only refresh the view.

Open document



Opens Document Pool with the filter for "Task" set to the current task, in the above image this happens to be "Cropping" for an example.

Fetch document



Opens the document with the highest priority with a "Work" status from current task. If there are no documents on Work status in the current task, no document will be opened. This button could be useful for large environments, where an operator does the same task, so they will no longer have to search for the next document

Project configuration



This will open the Control Center on "Configuration" tab. If Control Center is already open, pressing this button will switch the Control Center to the configuration tab.

Control Center



Opens Control Center on "Services status" tab. If Control Center is already opened, pressing the button will switch the tab to "Services status".

Help

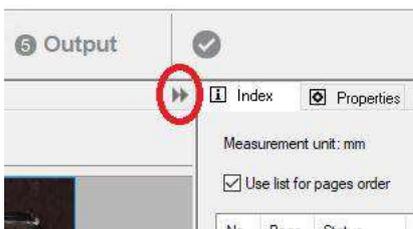


Opens the webpage with all the manuals.

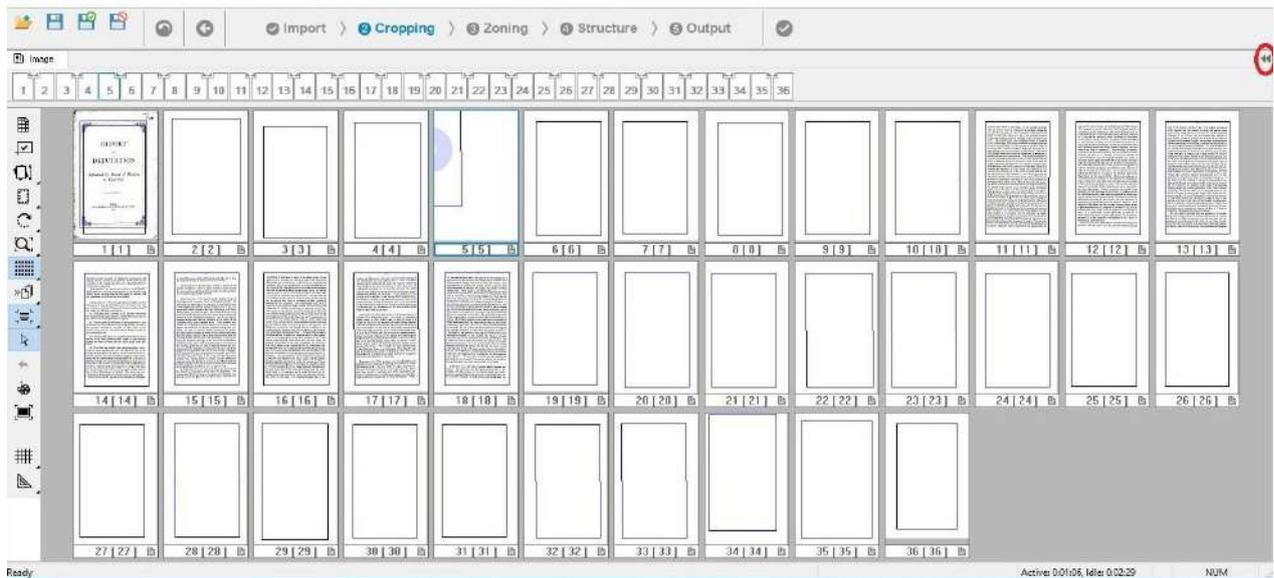
Working windows

The working windows on left and right hand side can be resized by dragging manually or by maximize/minimize buttons  or .

Example:



If  is pressed, the current view expands and fills the whole display area, but the workflow bar and page bar are still available:

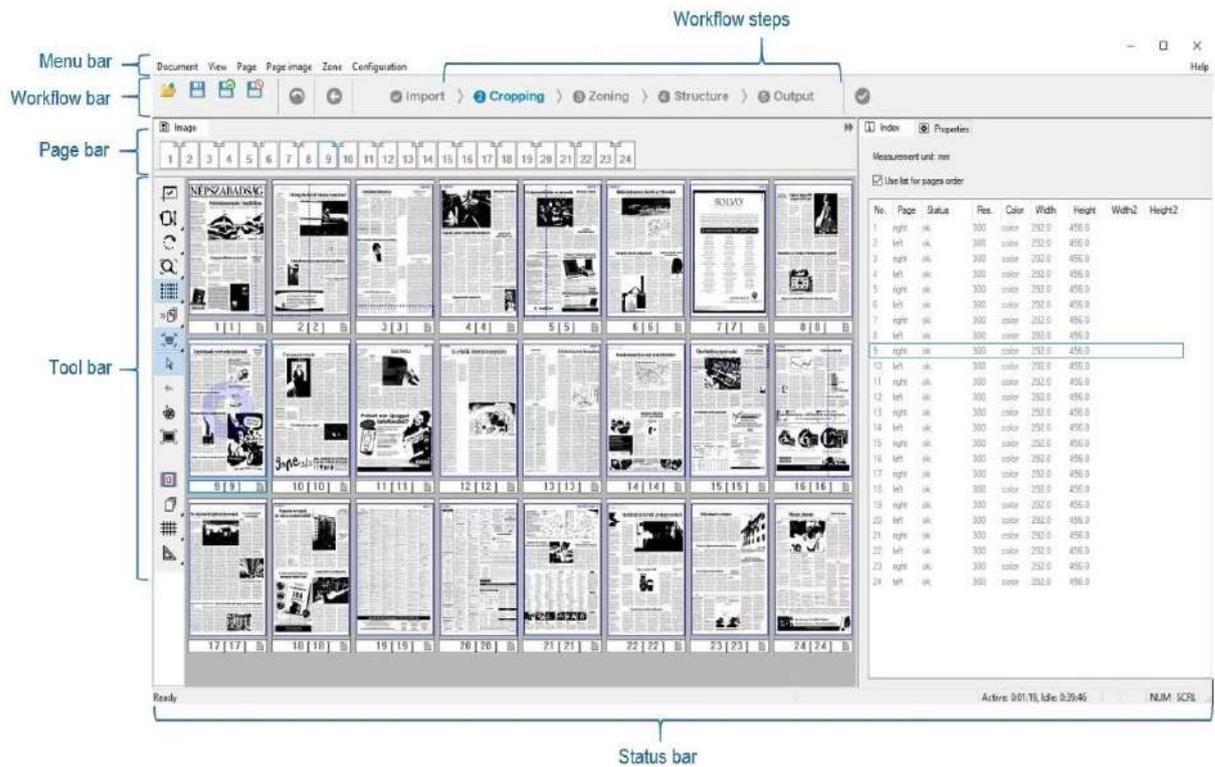


The option to maximize the view is not available for all docWizz views – Metadata or Clip view don't have the option to be expanded because for correction in these steps, both the left and right hand views are needed.

The selected view will be kept (user dependent) when docWizz is closed and reopened.

3.1 Bars

Overview of different bars in docWizz:



3.1.1 Workflow bar

In the workflow bar the processing steps and tasks are indicated. Depending on the current selected profile the amount and kind of steps may differ.

When opening a document, the workflow changes automatically to the step and task that the document is currently in. For example, when opening a document in **Cropping** the current task will be **Prepare cropping**. The previous task for the **Import** step is finished and has got a check mark.



Open

Open an existing document.



Save

Save the current document.



Set status

Set status, label, priority and save the document. Saves and closes the document.

Changes Status and/or Labels for selected documents ? X

Status:

Label:

Priority:

Enter your comment here:

Comments history:

Comment	Comment Date	Visible	User

Show all

Second method is to go to Document Pool and select one or multiple documents. Click then on

and the Label dialog opens from there.

Create a new label or select an existing one and click OK. To remove the label you select the >Empty< value and press the OK button.

Use the button to add new labels or remove existing ones.

User Labels ? X

- Intrerrupted
- Questions
-
- Verify Layout Zones
- Verify Page Numbers
- Verify Structure
- Verify OCR
- Final QA
- Red
- Green
- Blue

Reset on route

Label:

Comments:

If "Reset on route" is checked, the label of the document will be removed before routing the document to another step.

In the Document Pool you can filter by label. Only documents with the selected label will then be shown in the list.

Label:

-
-



Discard

Discard changes made in the current document.



Rescan

In the Rescan task bad scanned pages can be replaced or missing pages can be added.



Route document back to a previous step.

The dialog box is titled "Route document back to 'Review page sequence'". It features a dropdown menu at the top with the text "Keep working on this document in defined task.". Below this are fields for "Status:" (set to "Work"), "Label:" (set to ">Empty<"), and "Priority:" (set to "0"). There is an "Enter your comment here:" text area and a "Comments history:" table with columns for "Comment", "Comment Date", "Visible", and "User". The dialog includes "OK", "Cancel", "Add", "Copy", "Remove", and "Show all" buttons.

Process

To process the document to the next step you have 3 choices:

Use the Process button.

Click on the next step in the workflow bar.

In the Open Document window, use the "Route" button to select a task to route the document to.

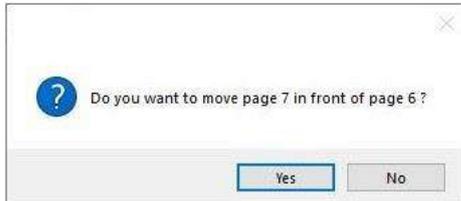
See [Process documents](#) chapter for details.

The dialog box is titled "Start processing". It features a dropdown menu at the top with the text "Process document now in user interface (dW Client)". Below this is a checked checkbox "Test rejects before routing." with a sub-note "Uncheck to compute rejects in background processing.". There are fields for "Status:" (set to "Work"), "Label:" (set to ">Empty<"), and "Priority:" (set to "0"). There is an "Enter your comment here:" text area and a "Comments history:" table with columns for "Comment", "Comment Date", "Visible", and "User". The dialog includes "OK", "Cancel", "Add", "Copy", "Remove", and "Show all" buttons.

Move pages

Note: This is not usually available past the Review Zoning task.

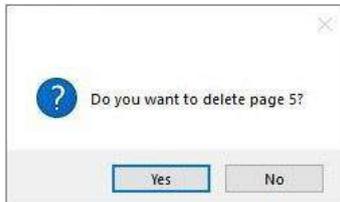
The order of pages in the page bar corresponds to the order in which they were scanned. You can change the order of the pages by moving the cursor to the page symbol and clicking and holding down the left mouse button. As soon as you move the mouse the cursor changes its appearance  and you can switch the location of the page in question to a new location. Before finalizing this operation the system will ask for confirmation of this task.



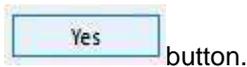
Before processing, the system prompts you to confirm the dedicated move.

Delete pages

Only available if image view in the left working window is selected.



Pages can be deleted by placing the cursor on the according symbol page in the page bar and by pressing the right mouse button. Select Delete from the context menu. The page will be deleted after you have answered the the prompt by pressing the

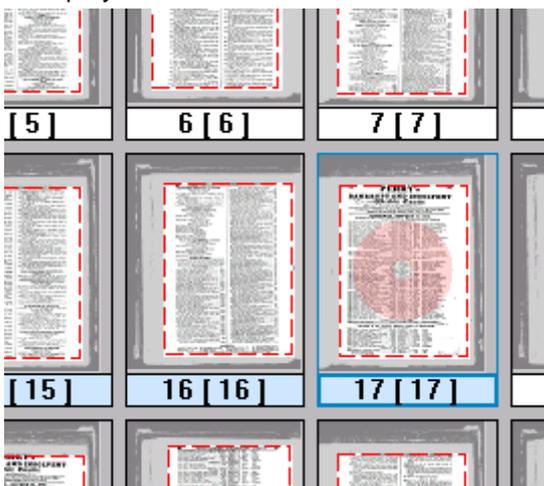


Select / Deselect pages

Use (Ctrl) or (Shift) to select several pages. Use (Space) to deselect all pages.

Dynamic page view  shows all selected pages and the currently clicked page.

In multiple page views the selection of pages work similar from the bottom line where the page number is displayed.

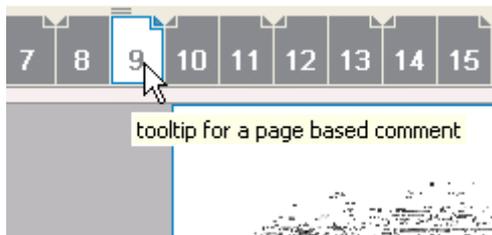


Page bar markers



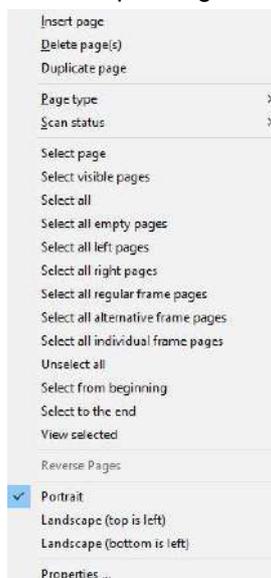
-  ok
-  bad OCR quality, bad resolution, wrong page and other errors
-  target
-  retained
-  missing
-  missing in original
-  as in original, page skewed in original, text cut off in original, too close to binding in original
-  not cut properly
-  double frame on image

Two small stripes on top of a page icon shows pages with comments (comments that have been made at page level in ScanClient or manually here in the properties interface).



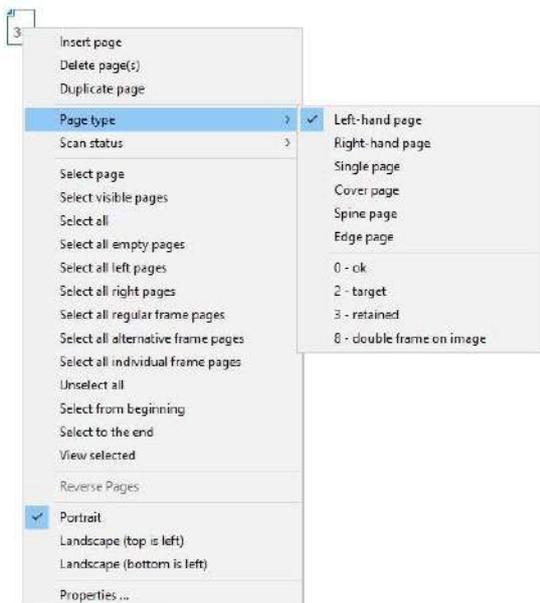
Page bar context menu

There is a context menu for each page as well, offering some more options. The content of this context menu is depending on the selected step.

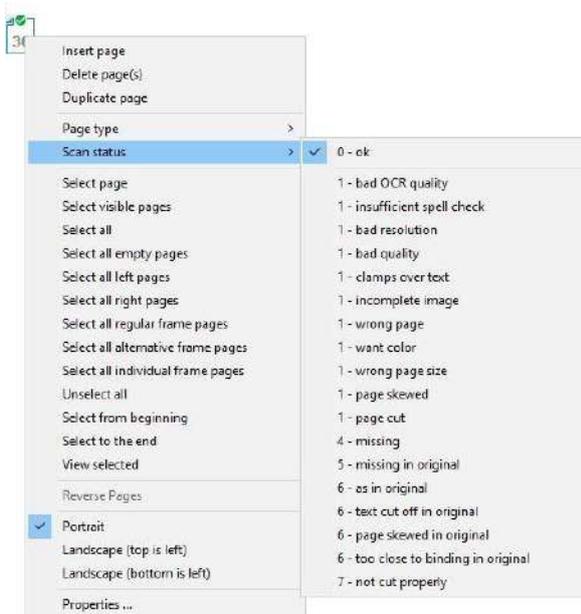


- **Insert Page:** Inserts a blank page in front of the current page.
- **Delete Page(s):** Deletes the current page.
- **Duplicate Page:** Duplicate the selected page.

- **Page type:**



- **Left/Right hand page:** Indicates whether the page is left hand or right hand. Important for reprinting purposes and adding adequate margins.
- **Single page / Cover page / Spine page / Edge page:** Also important for reprinting purposes: Tells in a double page book that the current page is a single page. So it will not be split nor handled as left or right hand page.
- **Scan status:**



- **Select Page:** Select single page: Selected page is marked in color. If you want to select several pages, activate the **Select Page** entry, hold the (Ctrl) key down and select one page after the other. Or use (Shift) to select several pages. In the next step, all selected pages are processed the same way. Example: rotate - all selected pages should rotate, all others should not. Use (Space) to deselect a page.
- **Select visible Pages:** Select only the pages that are in the image view, changing the view will select different number of pages.

- **Select All:** Selects pages all at once. Currently displayed page's page number is shown in black color.



- **Select all empty pages:** Select only pages that do not have frames.
- **Select all left/right pages:** Displays only the left/right side pages.
- **Select all regular frame pages:** Select only the pages with regular frames.
- **Select all alternative frame pages:** Select only the pages with alternative frames.
- **Select all individual frame pages:** Select only the pages with individual frames.
- **UnSelect All:** All pages are unselected.



- **Select from Beginning:** Selects pages from the beginning to the one that is currently selected.



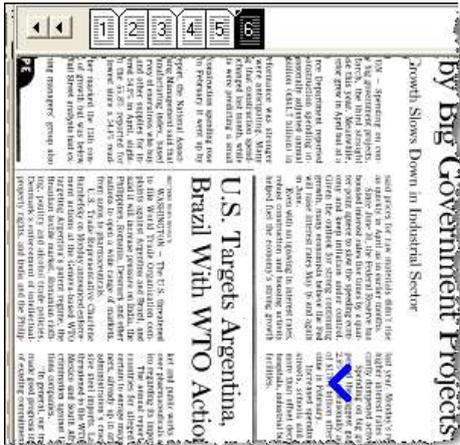
- **Select to End:** Selects pages from the selected page (page 9 in this example) to the end.



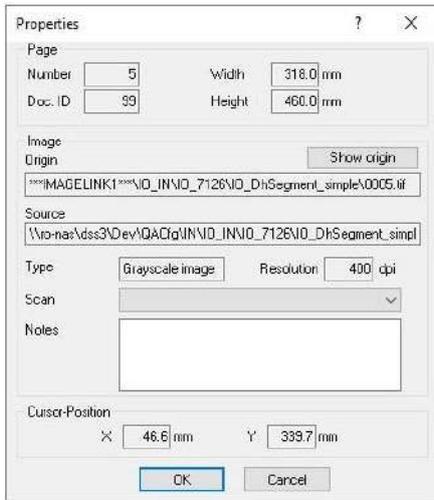
- **View selected:** Changes the view to show all selected pages.
- **Reverse pages:** Reverse the order of the selected pages.
- **Portrait:** The page orientation is portrait.
- **Landscape (Top is left):** In image view, the page orientation can be flagged. Right click on a page icon and select the orientation. If the orientation is not portrait, you will see a blue triangle on the side of the image which is marked as top. When exported the image will either be rotated or the orientation tag will be set.



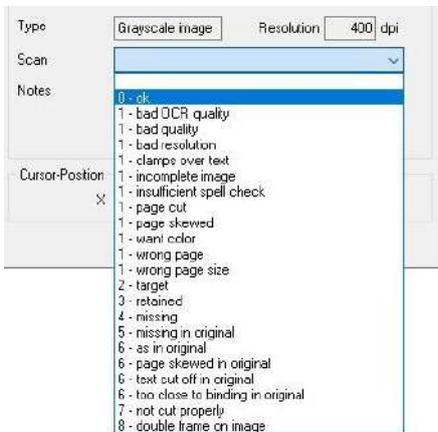
- **Landscape (Bottom is left)**



- **Properties:** Right click on an image to select Properties which for example resolution file name, source file destination and more.

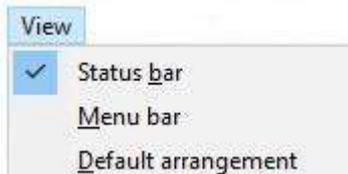


You can also specify here the scan quality and enter some notes.

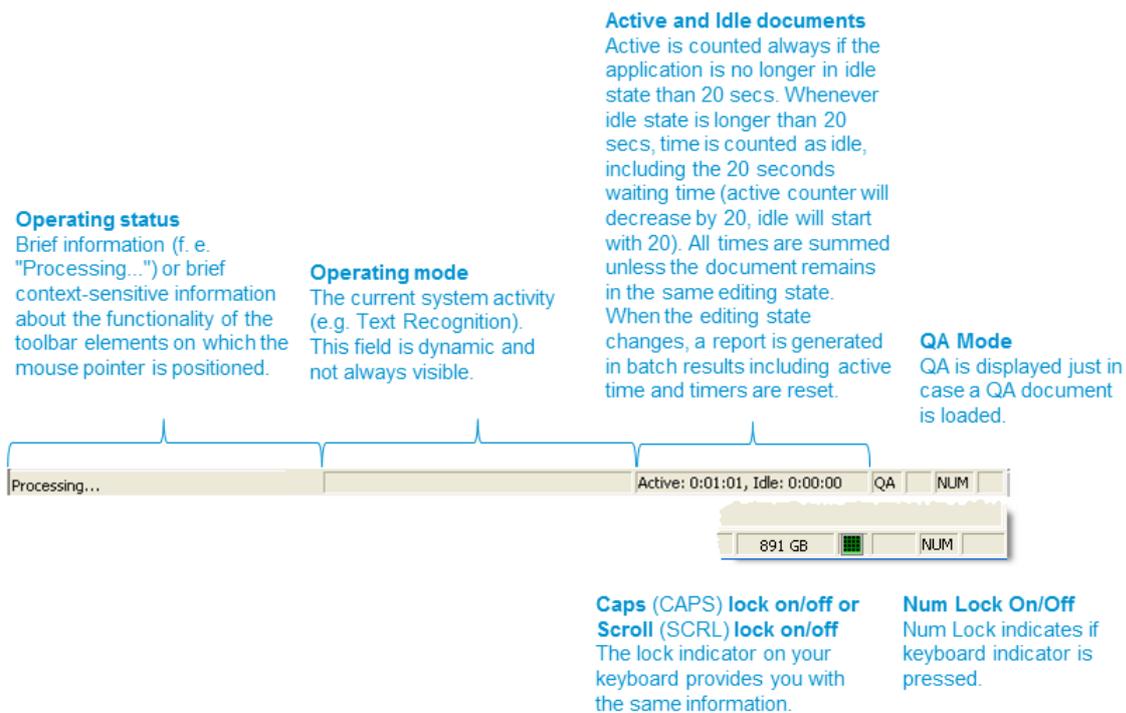


3.1.3 Status bar

Show the status bar permanently mark the "Status bar" entry in the View menu.



The status bar at the bottom of the docWizz window gives you additional information about the current and most recent tasks carried out. There are five sections providing information about the operating status of the system.



Num Lock On/Off Indicates whether the number lock for the numeric keypad on your keyboard is on or off. These keys have dual functions. They can be used either to input numbers, when the number lock is On, or when Off they function like arrow keys to move the cursor. Use the Num key to toggle back and forth. Num Lock indicates if keyboard indicator is pressed.

3.1.4 Menu bar

The menu bar contains lists of available menus and resides at the top.

The menu bar is hidden per default. To show it press (Alt) or (F10) key.

To show the menu bar permanently mark the "Menu bar" entry in the View menu.

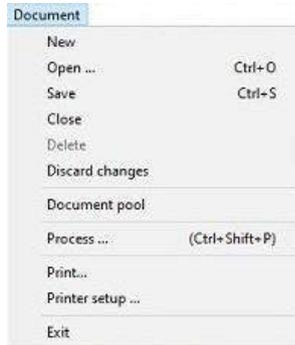
The combination, in which the menus are presented, always depends on the particular working situation. Not all functions of docWizz are always active. The status of a function depends on how your system is configured, in which program area you are currently in, and which program operations you have

executed. Available functions are displayed in black and deactivated functions are displayed in light gray. You can only execute active functions.

Depending on your operating system, the name of the menus and the single functions are displayed with a special character underlined. You can open the menu by pressing (Alt) key and the corresponding key. If the menu is already pulled down, the key performs the corresponding function.

3.1.4.1 Document menu

Document menu is available in all tasks.



New

Creating a new document requires the current task be changed to Review Import. You will be asked whether you want to store the current document and proceed to changing the task.

Open

Opens the Document pool window to manage any document in process. Shortcut: (Ctrl+O)

Save

Saves the document's current status. Shortcut: (Ctrl+S)

Close

Closes document. You will be asked whether you want to save the changes before closing.

Delete

Deletes document. Please confirm to delete it.



Discard changes

Discard changes made in the current document and closes the document.

Process

Process the document to the next step. See [Process documents](#) chapter for details.

Shortcut: (Ctrl+Shift+P)

Print, Printer setup

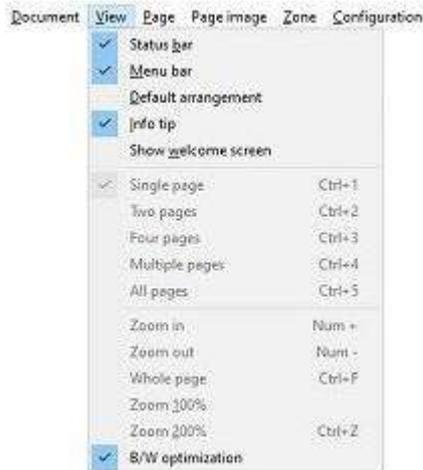
Prints the whole document on the selected printer. Use Printer setup to configure the printer.

Exit

Exit docWizz.

3.1.4.2 View menu

View menu is available in all tasks but there are different entries.



Status bar / Menu bar

Toggles view of the status or menu bar.

Default arrangement

Restores the default display of the icon and toolbar.

Info tip

By default the entry is checked. If unchecked, the info box will not be displayed when opening a document, but still available when hovering over the task. If unchecked, on restart the entry will be checked again.

Show welcome screen

Shows the welcome screen if checked.

Single Page

Displays only one page of the document in the working window. Shortcut: (Ctrl+1).

Two Pages

Displays two adjacent pages of the document in the working window – similar to an opened book. Shortcut: (Ctrl+2).

Four Pages

Displays two rows of pages of the document in the working window at the same time. Shortcut: (Ctrl+3).

Multiple Pages

Displays multiple pages of the document in the working window at the same time. Shortcut: (Ctrl+4).

All Pages

Displays all pages of the document in the working window. To achieve this, the pages have to be minimized considerably. Shortcut: (Ctrl+5). Maximum amount of pages shown is 50 + x (x means the complete last row is filled with preview pages, even if there are then more than 50).

Zoom In

Enlarges the image by increments. The (+) key on the numeric keypad performs the same function. Zoom-In is possible to pixel size (one image pixel is one screen pixel).

Zoom Out

Reduces the image by increments. The (-) key on the numeric keypad performs the same function. Zoom-out is possible to minimum 32 screen pixel width and height.

Whole Page

Resets the zoom and shows again the entire image page. The image is sized to the dimensions of the working window in such a way that its greatest possible spreading "usually vertically" fits into the working window.

Zoom 100%

1:1 view of the image. One pixel in the file corresponds to one pixel on the screen.

Zoom 200%

2:1 view of the image. One pixel in the file corresponds to four pixel on the screen.

B/W optimization

Optimizes the black-and-white display of the source page to improve viewing. This function is used to enhance legibility of scanned pages on the screen.

Next item

Calls up the next entry in the tree structure of the document.

Previous item

Calls up the previous entry in the tree structure of the document.

Select all of same class

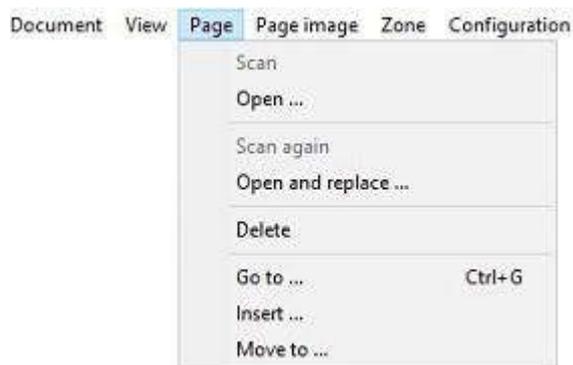
Selects all documents belonging to the same class of documents.

Collapse all

Collapses all branches of the tree structure.

3.1.4.3 Page menu

Page menu is available in the tasks: Setup Import, Review Import, Prepare Cropping and Review Cropping.



Scan

Scans the page that is currently in the scanner. The page is added to the pages already scanned. Use the Scan button to capture source documents by activating your system's installed and configured scanner.

Open

Opens a page image saved as a file. The page is added to the pages that are already open.

Scan again

Repeats the scanning process. The current page is replaced.

Open and replace

Opens another page image that replaces the current page.

Delete

Deleted the page marked as active from the document batch.

Go to ...

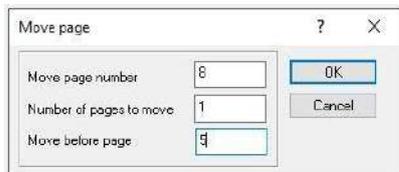
Helps you to move from one open page to another within the open page batch. This function can also be accessed from the page bar where the open pages are displayed. Shortcut (Ctrl+G).

Insert



Inserts one or more pages.

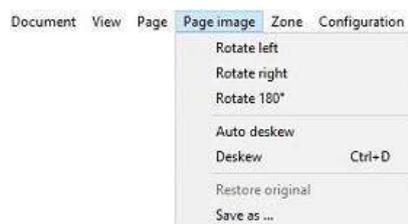
Move to ...



Move page from one place to another.

3.1.4.4 Page image menu

Page image menu is available in the tasks: Setup Import, Review Import, Prepare Cropping and Review Cropping.



Rotate Left

Rotates the page image 90 degrees to the left.

Rotate Right

Rotates the page image 90 degrees to the right.

Rotate 180°

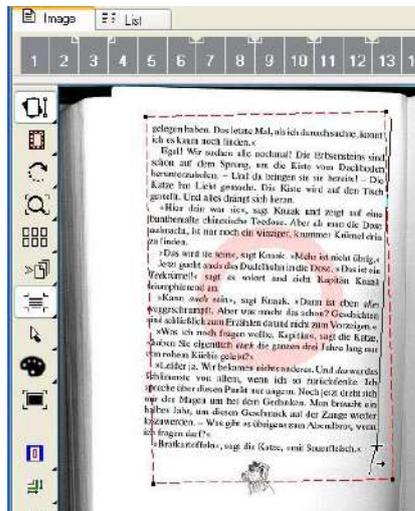
Rotates the page image 180 degrees, turning it upside down.

Deskew

Straightens crookedly scanned images of documents automatically. Using this function, the system automatically detects the right angle of the image, which is then rotated to a straight position. The effective use of this function depends on the appearance of the image and the scan quality. In some cases this does not lead to the expected result. If this is the case use the interactive function Deskew manually.

Deskew manually

Allow you to manually straighten crookedly scanned images of documents. After activating the function, the cursor changes its appearance. Put the cross of the cursor at a known horizontal or vertical line on the image such as a borderline, line between columns, type line etc. and draw a line by pressing down the left mouse button. As soon as you release the mouse button, the system rotates the image until the marked line is in the horizontal or vertical position.



Use this function only to correct slightly crooked documents. Using the deskew function to correct significantly crooked documents diminishes the optical quality of the characters, can irritate the OCR-process and lead to a bad quality of the recognized text.

If scanning has significantly crooked pages, it is recommended that they be scanned again accurately for further processing.

The manual deskew line must be at least 10 percent of the width/height of the image to be considered a valid deskew. This is to prevent accidental deskew.

Restore original

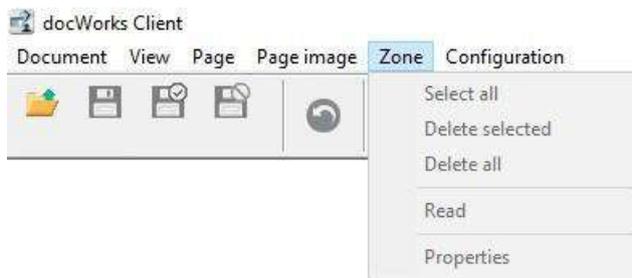
Restores the originally scanned image and undoes the actions performed on it.

Save as

As in a normal Windows environment, this function opens a selection window from which you can search for and select a folder to store the current page image as *.tif file.

3.1.4.5 Zone menu

Zone menu is available in Prepare Cropping and Review Cropping tasks.



Select All

Marks all the zones on the currently active page. If the zones are already marked, this function removes all the markings. Marked zones are highlighted in color.

Delete Selected

Deletes the selected zones on the currently active target page. The (Del) key also performs this function.

Delete All

Deletes all zones on the currently active target page.

Read

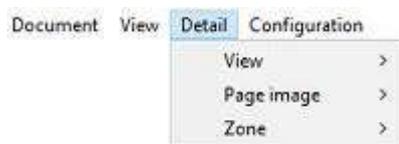
Interprets the contents of the currently selected zone and displays the result.

Properties

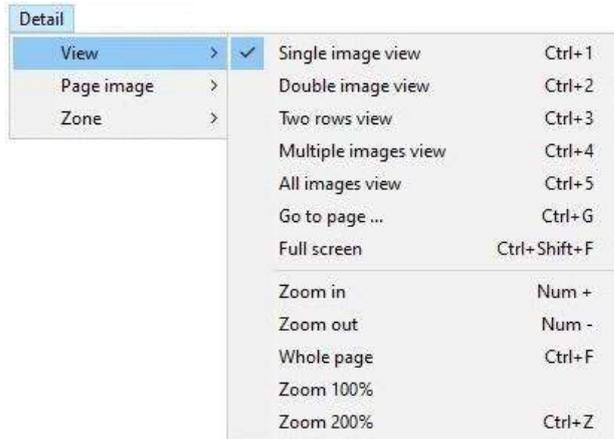
Displays a dialog that shows you the properties of the zone currently selected.

3.1.4.6 Detail menu

The Detail menu shows three separate sections: View, Page image and Zone and is available in the following tasks: Review zoning, Review page sequence, Review issues, Review structure and Text, and Review output.



View



Single image view: Displays only one page of the document in the working window. Shortcut: (Ctrl+1)

Double image view: Displays two adjacent pages of the document in the working window – similar to an opened book. Shortcut: (Ctrl+2)

Two rows view: Displays two rows of pages of the document in the working window at the same time. Shortcut: (Ctrl+3)

Multiple image view: Displays multiple pages of the document in the working window at the same time. Shortcut: (Ctrl+4)

All image view: Displays all pages of the document in the working window. To achieve this, the pages have to be minimized considerably. Shortcut: (Ctrl+5). Maximum amount of pages shown is 50 + x (x means the complete last row is filled with preview pages, even if there are then more than 50).

Go to page ... : Go to a certain page number. Shortcut: (Ctrl+G)

Full screen: You can enlarge the active window to use the full width of the screen. Click  button to go back to default view.

Zoom In: Enlarges the image by increments. The (+) key on the numeric keypad performs the same function. Zoom-In is possible to pixel size (one image pixel is one screen pixel).

Zoom Out: Reduces the image by increments. The (-) key on the numeric keypad performs the same function. Zoom-Out is possible to minimum 32 screen pixel width and height.

Whole page: Resets the zoom and shows again the entire image page. The image is sized to the dimensions of the working window in such a way that its greatest possible spreading "usually vertically" fits into the working window.

Zoom 100%: 1:1 view of the image. One pixel in the file corresponds to one pixel on the screen.

Zoom 200%: 2:1 view of the image. One pixel in the file corresponds to four pixel on the screen.

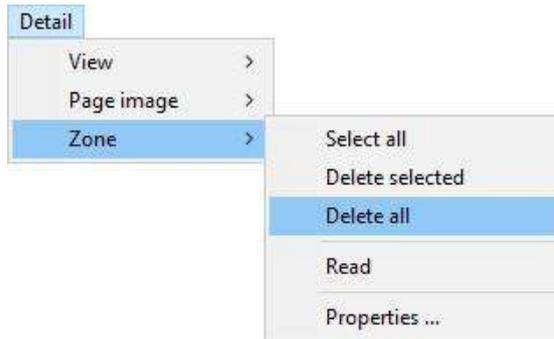
Page image



Restore Original: Restores the original image and undoes the actions performed on it.

Save as ... : Saves the original page of the current page image. This command opens the Windows window for saving files. By using this you can select the folder the image file is to be saved in.

Zone



Select All: Marks all the zones on the currently active page. If the zones are already marked, this function removes all the markings. Marked zones are highlighted in color.

Delete Selected: Deletes the selected zones on the currently active target page. The (Del) key also performs this function.

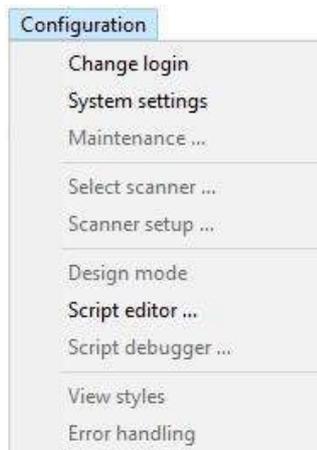
Delete All: Deletes all zones on the currently active target page.

Read: Interprets the contents of the currently selected zone and displays the result.

Properties ... : Displays a dialog that shows you the properties of the zone currently selected.

3.1.4.7 Configuration menu

The **Configuration menu** is available in all tasks and provides various functions that you can use to make specific system settings. As a rule, some menu items are password-protected, depending on the installation. Access to these functions is regulated by the login name and password you enter at start-up time. If you are not sure whether you have access, call your system administrator.



Most of the functions in the **Configuration menu** are only accessible to administrators or to users with administrator permission. For other users, these functions become inactive and appear gray.

The commands in the Configuration menu have the following functions:

Change Login

Allows another user (the administrator, for example) to log in during a docWizz session without requiring the current user to exit the program.

System settings

Presents a menu for making system configuration settings. Further information is available in the docWizz Reference book.

Script editor ...

You need administrator permission (login and password) to enter this area.

The execution of the program functions can be influenced with the help of the script editor, where scripts can be written, edited and test-run.

Further information is available in the **docWizz Reference book**.

3.1.4.8 Help menu

The Help menu is available in all tasks. With the entries in the Help menu you can open the documentation files, see the error log and statistics and check the registration.

PDF Documentation

This function enables you to refer to the PDF documentations of docWizz.

Online Documentation

Opens the web page with all the manuals.

Error Log

This function enables you to refer to the Error Log window that automatically lists any errors that have occurred during the current docWizz session. In this way, support staff and docWizz administrators have optimal support when looking for the cause of irregularities in the running of the program.

Error dialog shows Code and Context as additional columns. Script command reporterror has optional parameter errorcode, which will be reported in error log database.

Column "Code" will contain an unique error code to identify which error has occurred. New column "Context" contains information like document ID. This helps to select all errors related to a single document. Databases will be extended on first start automatically.

The context menu (available on right click) contains, besides the "Copy cell" and "Copy row" options, a new option: "View more". This action will open a new dialog with the details of the selected comment / error.

The screenshot shows the 'Error log' window with a table of error entries. A context menu is open over the entry with ID 53445, showing options: 'View more...', 'Copy cell', and 'Copy row'. The 'View more...' option is selected, opening an 'Error details' dialog box.

ID	Date	Type	Message	Computer name
53453	2016/09/12 16:21:34	8	XMLDOCLOAD\Bogdan3D\docwork_69\c\g\PROJECT-CFG\AAA...	BOGDAN3DWC...
53452	2016/09/12 16:20:59	8	XMLDOCLOAD\Bogdan3D\docwork_69\c\g\PROJECT-CFG\AAA...	BOGDAN3DWC...
53451	2016/09/12 16:08:19	8	XMLDOCLOAD\Bogdan3D\docwork_69\c\g\PROJECT-CFG\AAA...	BOGDAN3DWC...
53450	2016/09/12 14:00:18	8	XMLDOCLOAD\Bogdan3D\docwork_69\c\g\PROJECT-CFG\AAA...	BOGDAN3DWC...
53449	2016/09/12 13:59:56	8	XMLDOCLOAD\Bogdan3D\docwork_69\c\g\PROJECT-CFG\AAA...	BOGDAN3DWC...
53448	2016/09/12 13:48:50	8	Script Work Task execution failed: source [expandpath ""DATA""\SC...	BOGDAN3DWS...
53447	2016/09/12 12:48:49	8	Script Work Task execution failed: source [expandpath ""DATA""\SC...	BOGDAN3DWS...
53446	2016/09/12 12:41:39	8	Script source [expandpath ""DATA...	
53445	2016/09/12 12:41:39	8	The same or similar error appeared 3 times:	
53444	2016/09/12 12:04:26	8	ies 14	
53443	2016/09/12 10:48:27	8	ies 14	
53442	2016/09/12 10:08:56	8	\doc	
53441	2016/09/12 10:08:47	8	\doc	
53440	2016/09/12 10:07:54	8	XMLDOCLOAD\Bogdan3D\doc	
53439	2016/09/12 10:07:39	8	XMLDOCLOAD\Bogdan3D\doc	
53438	2016/09/12 09:48:44	8	Script Work Task execution failed:	
53437	2016/09/12 09:48:39	8	Script Work Task execution failed:	
53436	2016/09/12 01:48:31	8	Script Work Task execution failed:	

Error details

ID: 53445
Date: 2016/09/12 12:41:39
Type: 8
Computer name: BOGDAN3DWSrv2
Error code: 103001
Context: PkID: 9140

Error message:
The same or similar error appeared 3 times:
Script Work Task execution failed :
source [expandpath ""DATA""\SCRIPT\Task\createPriorityEmpFileForFtp.tcl]
CreateFTPPriorityFile
The result was "invalid command name ""
while executing
"SrcConfigFile -getrootnode"
(procedure call: CreateFTPPriorityFile -line 14)
invoked from within
"CreateFTPPriorityFile"
Context:
ID:103001

Statistics

This function enables you to refer to the **Processing Statistics** window, which gives you details about the total count of processed pages, users, tasks, document types or projects. Please log in as administrator.

OCR usage statistics

To check the OCR consumption, the users can use the "OCR usage statistics" dialog from Help menu. The dialog is available for Admin users only. For more details, please, check in the Admin manual "Help menu" chapter.

About docWizz Client

This function enables you to refer to the information window, which gives you details of the version of docWizz installed on your system.

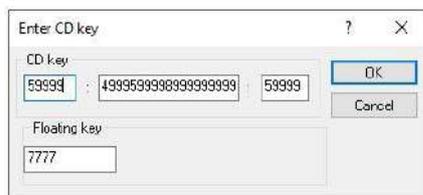
Here you can see the name of the User who is currently working on the system, which CD Key you are using, and also the Registration-ID of the machine.

You can access different system components. This depends on the CD Key you have received from CCS. You can retrieve information about your module access rights from a system list, using the button.

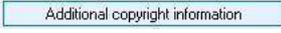
If you would like to extend your system with new modules or features or even additional licenses, these can be activated with a new CD Key, Click the  button and enter your new CD Key number. This is restricted to administrators.

Enter the floating license into the field "Floating key". This stores the possible number of parallel docWizz instances. This key comes from CCS.

The dialog ensures and validates, that the entered codes are valid and stored in the right ini (custom-glbl ini).



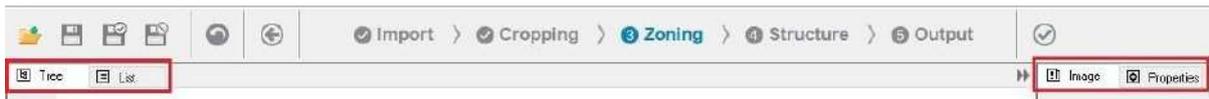
Complete the entry by clicking the (OK) button.

See additional copyright information by clicking the  button.

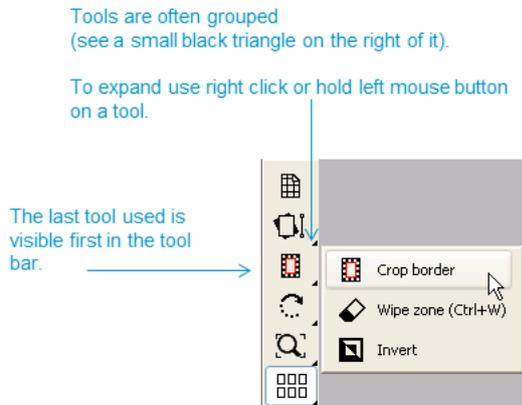
A selection of a component shows the additional copyright information.

3.2 Tools in different views

Tool bars serve as an always-available, easy-to-use interface for performing common functions.



Each processing step needs a different view to work on. An operator can switch between views on the left and on the right hand side to find the best working method for the current document. In some interactive steps there not all views are available, because they may not make sense for that specific step.



It is configurable if the toolbar should expand automatically. See **Expand tools automatically** chapter, in **docWizz ReferenceBook**.

3.2.1 Image view

Image view tools for the left or right hand side view depend on the task and not all are available in any task.



Open File

The Open button opens previously scanned and saved files. The standard Windows "Open File" dialog box displays the files from which you make your selection. You can select multiple files at once.

Direct import of PDF documents is supported. The text inside the PDF instead of OCR might be used here.



Split Document

Opens the Virtual Printer files dialog box. You can select and open a document for further processing.



Open

Divides a page image into two parts, if you have double pages.

Splits large batches of huge images that can't be processed as one single document in dW, especially when layout analysis and OCR is required. For that reason, those batches need to be split in the Review import task.

Shortcut: (Ctrl+T)

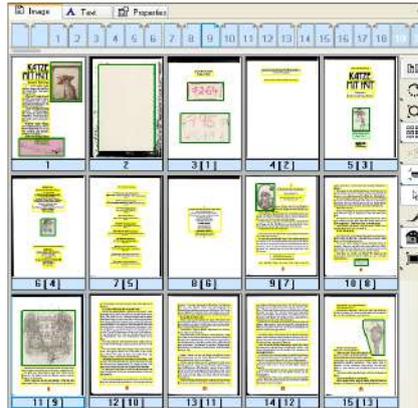


Show page grid

Shows a grid on all pages. Shortcut: (G)



Logical Page Numbers Displays logical page numbers in the page bar:



Default type

Click on this tool, if you want to create a new zone type, for example: Table, Textblock, Advertisement, Formula, Illustration or Vertical Textblock.

Sequence of zone types can only be changed, if a user is logged in having dW SYSTEMCFG rights. Default Admin user has those rights.

Select default type in left hand area and click . Then select the zone type on the right hand side, use drag and drop to place in the desired position.



Shrink text enabled

The button is available in normal and in full screen mode and in all steps. The operator can decide whether he wants dW crop the zone and the zone snaps to the text automatically or let the zones as they are.



Merge with next page
"Stitching"

This is a special function used for "stitching" - used for semi-automated creation of complete page images. It is available when configured. Stitching is used to stitch two half scanned pages into a single one. Used for special customer projects. This feature is available in the "Review import" task. See [Merge \(Stitching\)](#) chapter.



Deskew manually

Allows you to manually straighten crookedly scanned images of documents. After activating the function, the cursor changes its appearance. Put the cross of the cursor on a significant part of the image (borderline, line between columns, type line etc.) and draw a line by pressing the left mouse button and drawing along a known vertical or horizontal line in the image. As soon as you release the mouse button, the system rotates the frame until the marked line is in horizontal or vertical position.

Use this function only to correct slightly crooked documents. Using the deskew function to correct significantly crooked documents diminishes the optical quality of the characters, and can irritate the OCR-process which may lead to bad OCR quality.

Note: If document has a significant number of crooked pages, it is recommended that they be scanned again accurately for further processing. The manual deskew line must be at least 10 percent of the width/height of the image to be considered a valid deskew. This is to prevent accidental deskew.



Auto deskew

Straightens crookedly scanned images of documents automatically. Using this function, the system automatically detects the right angle of the image, which is then rotated to a straight position. The effective use of this function depends on the appearance of the image and the scan quality. In some cases this does not lead to the expected result. If this is the case use the interactive function **Deskew manually**.

Note: In Prepare cropping automatic deskew actions are applied on frames. If multiple pages are selected, the actions are applied on selection. In other tasks, the actions are applied on pages.



Deskew to the left/right

Straightens crookedly scanned images of documents slightly to the left/right.



Crop Border

Not available in Prepare cropping task. Everything outside the red frame is cut. The red frames appear in the Prepare cropping task.



Wipe Zone

Not available in Prepare cropping task. The **Wipe Zone** button helps you to deselect a zone that is not required for archiving (one which would possibly cause problems for text recognition).

Select the Wipe Zone tool. Cursor symbol changes to . Draw a rectangle around the area you want to wipe out. Click once with left mouse button inside the area to hide the content.



Invert

Not available in Prepare cropping task. You can use the **Invert** tool to invert an active zone and reverse the tone values. This means that black is converted to white and vice-versa. You might want to use this option when you have a source document with areas with white text on a black background, used mainly to emphasize a passage. These zones cannot be processed for automatic text recognition unless they are inverted first. Note: OCR text recognition in general is only working on black characters.



Rotate right

Rotates the page image 90 degrees to the right.



Rotate left

Rotates the page image 90 degrees to the left.



Rotate 180°

Rotates the page image 180 degrees, turning it upside down.

Note: In Prepare cropping the rotate actions are applied to frames. If multiple pages are selected, the actions are applied on selection. In the other tasks, the actions are applied to pages.



Zoom in

Enlarges the image by increments. The (+) key on the numeric keypad performs the same function. Zoom-In is possible to pixel size (one image pixel is one screen pixel).



Zoom out

Reduces the image by increments. The (-) key on the numeric keypad performs the same function. Zoom-Out is possible to minimum 32 screen pixel width and height.



Lock zoom

If a zoom factor is selected for one page and icon is pressed, all other pages are shown in the same zoom level (enlarged or reduced).



Display Entire Page

Resets the zoom and shows again the entire image page. The image is sized to the dimensions of the working window in such a way that its greatest possible spreading "usually vertically" fits into the working window. Shortcut: (Ctrl+P)

Zoom by right mouse key

Use right mouse button and drag a rectangle inside the page view to zoom once. To zoom to next levels use (Ctrl) key and right mouse button and drag again a rectangle around the desired area.

Cancel zoom by the  **Display Entire Page** button.



Lock view

Click on this icon to lock the current view of the right window. This is helpful if you want to compare for example table of contents on the right hand side with tree view on the left hand side. You can release the lock by clicking once again on this icon. This option, was introduced to keep the view in place when correcting OCR. Without it, the word rectangle is displayed for each word, with the Lock view option, the whole zone is displayed, it is mostly used in OCR correction and in list correction.



Magnify window

Move mouse cursor (without any click) over the original image, the magnification is shown parallel in the **Magnify** window.

We recommend to use this functionality with full screen mode.



Zoom in / Zoom out



Refresh image.



Durably fix a complete page. This is needed for example to compare different pages with a table of contents.

Operation: Press button - page remains, the cutout part will not be changed by moving the cursor over the original image. Clicks in image view change the cutout only, if these are executed on the fixed page. To fix on another page this must be selected first in image view and then click in the Magnifier on "Refresh".



Low resolution image view.



High resolution / colored view.



Gray scale view. On page change in Image view the magnifier is redrawn with the same coordinates, but with the new image.



No change of image tool, the image keeps unchanged.



Single image view

Displays only one page of the document in the working window. Shortcut: (Ctrl+1)



Double image view

Displays two adjacent pages of the document in the working window – similar to an opened book. Shortcut: (Ctrl+2)



Two rows view

Displays two rows of pages of the document in the working window at the same time. Shortcut: (Ctrl+3)



Multiple image view

Displays multiple pages of the document in the working window at the same time. Shortcut: (Ctrl+4)



All images view

Displays all pages of the document in the working window. To achieve this, the pages have to be minimized considerably.

Maximum amount of pages shown is 50 + x

(x means the complete last row is filled with preview pages, even if there are then more than 50). Shortcut: (Ctrl+5)

	Dynamic view	Show pages as needed. Could be helpful for Review Issues, where you can see only pages belonging for example to the front part or to the main part. Shortcut: (Ctrl+0)
	Previous page	This button is only active if you have selected  Display single page before. The view of the page jumps to previous page.
	Next page	This button is only active if you have selected  Display single page before. The view of the page jumps to next page.
	Shrink to text	The button is available in normal and in full screen mode and in all steps. The operator can decide whether he wants dW to crop the zone and then snaps to the text automatically or leave the zones as they are.
	Properties	Opens the property dialog (in full screen mode only).
	Default Zone Tool	Activates the default tool to create zones and select them. You can also press the (Esc) key to call up this tool.
	Undo	Reverses the last operation that you have performed on a source page in the working window. It restores all changes to the frame, no matter the action used (click, click and drag, drag new frame, move frame etc.).
	Brightness	<p>You can use the Brightness button to change the brightness of the image – especially for use in OCR-processing. This function is shown as active only when your document has been captured and saved as a grayscale image.</p> <p>Pressing the Button opens the Brightness dialog box. You can use the slider to adjust the brightness of the image and to assess the effects on the screen in real time. When you are satisfied with the results, you can store the source page.</p> <p>Moving the slider to the left or to the right moves the threshold – in a range of 256 gray values – up or down, at which values of gray are converted to either white or black. The threshold is displayed as a number in the window on the right side of the slider. If you have applied Zones to the current page all changes will only be performed on zones that are activated.</p> <p>If you check the Invert <input checked="" type="checkbox"/> box, this will reverse the brightness values of the page to produce a negative of it. This helps to distinguish better between the foreground and background.</p> <p>If you check the As Picture box, the marked zone will be treated as a graphic image and separation is made only on the basis of the brightness. If the box is not checked, the system looks for a background color and treats all elements, other than the background, as text.</p> <p>The Error Diffusion box depends on the As Picture check box and is checked to make the images of pictures on the source page appear better.</p>
	Image Zones gray	Toggles displaying image zones as gray image.
	Display color/grayscale images	Toggles between the display of images in grayscale/color or black-and-white. Usually page images are displayed in dW as black-and-white (1 Bit) images.
	Automatic brightness	You can use the Automatic Brightness button, which changes the brightness of the image automatically to suit your requirements – especially for use in OCR-processing.



Merge zones

Merges two or several selected zones to one single zone. You can also perform this function by pressing the (F8) key. All active zones of the current page will be merged.



Polygon zone

You can use the **Polygon zone** button, or press the (Ctrl+P) key, to create a polygonal (many-sided) zone for clipping a text column with a formed setting. Once you have activated the function, place the mouse pointer at the point you want the polygon to begin and click the left mouse button to indicate the starting point of the polygon. When you move the mouse a line appears on the screen beginning at the polygon's starting point and, behaving as if it were magnetically attracted to the mouse pointer, follows it, but without leaving the north-south and east-west orientation. As soon as you press the (Shift) key the line moves at any angle you want around the starting point and connects it directly with the mouse pointer. See [Knife - Polygon](#) chapter.



Knife

The **Knife** function is used to split zones. Select the function by clicking the **Knife** button or the (F9) key. The mouse pointer changes to a knife . Move the knife in the active zone to the position where you want to make the cut. Clicking the left mouse button with a slight movement of the mouse cuts the zone into two pieces. Exit the function by pressing the (K) or the (Esc) key.

A zone cut here is a normal cut, that cuts only when an empty area is found. See [Knife - Polygon](#) chapter.



Full screen

Displays page on the entire screen. Click on this icon again to return to normal user interface. Shortcut: (Ctrl+Shift+F).

Pressing the Full zoom button again restores the normal view. The zoom function can also be used in the Full zoom mode to enlarge the pages to be processed.



Fast correction of zones (zone editing)

Fast correction mode works in full screen mode only and on steps like Review Zoning or Page Sequence (steps higher than review Cropping). See [Fast correction of zones](#) chapter.

3.2.2 List view

List view tools:



Next item

Calls up the next entry in the tree structure of the document. Shortcut: (Ctrl+N).



Previous item

Calls up the previous entry in the tree structure of the document. Shortcut: (Ctrl+P).



Reject manager

Opens the rejects in a separate window and is now independent from the current step.

The reject manager window can be moved. See [Rejects](#) chapter.



Configuration Dialog

Opens configuration dialog. This functionality is only available for users with special administrator permissions.

	Configure error handling	Administrator login required.
	Refresh	<p>Refreshes the current list. Shortcut: (F5)</p> <p>Refresh will only re-populate the list with new items if any are existing. Refresh will re-run the reject script.</p> <p>By default accepted rejects are not re-calculated. If the reject is accepted, and the reject has been disabled or it should not appear, you will have to manually reject it, then refresh the list.</p>
	Compute properties	<p>Dynamically computes properties. This is not just for OCR computing - but for any other dynamic properties that need to be computed (OCR is the most used property that can be computed also in the list). Projects might have custom properties that can be computed. It can be done on any property that takes time to be computed.</p>
	Group mode	<p>Toggles grouping mode. Shortcut: (F3)</p> <p>You select a structure element (e.g. an article) in tree view and then you simply click on those zones (or drag a frame to cover more than one) that should belong to the article as well but are not contained yet. See Subtask: Review structure chapter.</p>

3.2.3 Text view

See [How to use the OCR text correction](#) for more details.

	Zoom in	<p>Enlarges the image by increments. The (+) key on the numeric keypad performs the same function. Zoom-In is possible to pixel size (one image pixel is one screen pixel).</p> <p>Shortcut: (Num. +)</p>
	Zoom out	<p>Reduces the image by increments. The (-) key on the numeric keypad performs the same function. Zoom-Out is possible to minimum 32 screen pixel width and height. Shortcut: (Num. -)</p>
	Lock zoom	<p>If a zoom factor is selected for one page and icon is pressed, all other pages are shown in the same zoom level (enlarged or reduced).</p>
	Fit width	<p>Resets the zoom and shows again the entire image page. The image is sized to the dimensions of the working window in such a way that its greatest possible spreading "usually vertically" fits into the working window. Shortcut: (Ctrl+P)</p>
	Zoom by right mouse key	<p>Use right mouse button and drag a rectangle inside the page view to zoom once. To zoom to next levels use (Ctrl) key and right mouse button and drag again a rectangle around the desired area. Cancel zoom by the  Display Entire Page button.</p>

	Lock zoom factor	If a zoom factor is selected for one page and icon is pressed, all other pages are shown in the same zoom level (enlarged or reduced).
	Display color/grayscale images	Toggles between the display of images in grayscale/color or black-and-white. Usually page images are displayed in docWizz as black-and-white (1 Bit) images.
	Lock view	Click on this icon to fix the current view of the right window. This is helpful if you want to compare for example table of contents on the right hand side with left tree view. You can release the fix by clicking once again in this icon. This option, was introduced to keep the view when correcting the OCR. Without it, the word rectangle is displayed for each word, with Fix View option, the whole zone is displayed, it is highly used in OCR Correction and in list correction.
	Undo	Reverses the last operation that you have performed.
	Previous error	Lets the cursor return to the previous error. Shortcut: (Shift +Tab)
	Next error	Lets the cursor jump to the next error. Shortcut: (Tab)
	Search	<p>Opens the Find dialog, that helps you to search for special words in the full text.</p> <p>Type the search string into the Find what: input field and specify conditions for the search using the check boxes.</p> <p>Inexact search is usually used for running titles or multiple paste.</p> <p>Clicking Inexact search, the other two check boxes are not grayed out they can be used in combination.</p>
	Replace	<p>Opens the Search and Replace dialog that helps you to search for special words and replace them with another.</p> <p>Find/replace with "inexact search" - usually used for running titles or multiple paste</p> <p>Use the little tick box on front of the found entries to click then "Replace".</p>



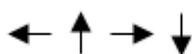
Error word list

Switches between the full text display and the list view.

In the **Error word List**, the system shows errors in a two-column list. The left column contains the system's color-coded OCR interpretation of the words, and the right column shows the original image of the text.

The bottom part of the text correction window, occupying the full width of the screen, shows the original section of the document that contains the text images shown in the above columns; this is helpful if you need to review the context of the words you are correcting.

The list shows only those characters, the system was unable to recognize with the specified certainty. The list view is particularly suitable for rapidly working through errors. Some of the functions in the list view are different from the full text view.



You can:

use the cursor keys to move quickly through the lines.

double-click on an unrecognized character to show the context of the character.

The toolbar buttons also retain their normal functions in list view mode.



View word

You can use this button to get information about the source of misinterpreted words quickly. To do so, place the cursor on the word and click the **Show word** button. The corresponding part of the document where the word originated then appears in the image window and is color highlighted

facilities. Shortcut: (Alt+V)



Accept

If a word has been marked as an error but is correct, click the **Accept** button to tell the program that the word is correct. The word then appears in black. It is not added to the dictionary. Shortcut: (Alt+1)



Add to dictionary

The **Add to Dictionary** button lets you add words to the dictionary that were previously unknown to the system. Shortcut: (Alt+L)



Correct automatically

The **Correct Automatically** button carries out the automatic correction of words throughout a document, for example repeated OCR errors, or words that were misspelled in the original – for example "Mitterand" instead of the correct "Miterrand". Select the incorrectly spelled word, correct it, and click the **Correct Automatically** button. Shortcut: (Alt+K)



Invalid word

Use the **Invalid Word** button to designate a word as invalid. The dictionaries accept complex words, which do not always necessarily make sense. For example: booklet "Book-Let" would be acceptable to the dictionary as the components of this complex word are acceptable, but the composite word would have to be declared as an invalid word. Such invalid words will not be accepted and appear in red. To undo such a designation you must edit the dictionary. Shortcut: (Alt+U)



Don't correct

You can use the **Don't Correct** button to specify those words that should not be automatically corrected, or which should be undone. Words are corrected automatically by text recognition. In some cases, corrections that you do not require. In this case, select the word in blue and click on the **Don't Correct** button. Then spell the word as you wish it to be spelt.

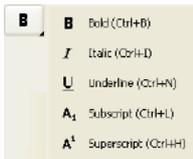


Dictionary

Press the **Dictionary** button to open the **User Dictionaries** dialog box.

This window has four sections and is used for managing the words system users have added to the dictionary using the following functions: **Add to Dictionary** (shown under **New Words**), **Correct Automatically** (shown under **Automatic Corrections**), **Invalid Word** (shown under **Invalid Words**), and **Don't Correct** (shown under **Do not Replace**). To delete words in these windows, simply select them with the mouse

and click the  recycle bin icon.



Change text into selected font shape.



Print

Print only the text on a selected printer.

3.2.4 Tree view

Tree view tools for the left hand side view:



Next item

Calls up the next entry in the tree structure of the document.



Previous item

Calls up the previous entry in the tree structure of the document.



Select all of same class

Selects all documents belonging to the same class of documents. Shortcut: (Ctrl+S).



Collapse all

Collapses all branches of the tree structure.



Level up

Moves the item up one level in the items hierarchy. Shortcut: (F6)



Level down

Moves the item down one level in the items hierarchy. Shortcut: (F7)



Reject manager

Opens the rejects in a separate window and is now independent from the current step.

The reject manager window can be moved. See [Rejects / Rejects manager](#) chapter.



Toggle error

Toggle the error status of the selected item.

	Reset Error	Resets the error status of the current item. Especially used for page number sequence checking, pages where the page numbers are out of sequence get a flag "Error". An operator may accept (reset, for example when indeed there is a new sequence started) or he may set again the error status in case he detects something wrong by his own.
	Set Error	Sets the error status of the current item. See description for Reset error.
	Previous error	Jumps directly to the previous item, the system has identified an error. To get this or the next button active, use the Go to (Ctrl+G) functionality and check there the Page only with error status entry. Shortcut: (Shift+page up)
	Next error	Jumps directly to the next item, the system has identified an error. These error tools are available in the tasks Review Zoning, Review Page Sequence, Review Issues or Review Structure and Text. Shortcut: (Shift+page down)
	Configure error handling	Administrator login required.
	Update properties	Dynamically computes properties (OCR)
	Go to page	Using the Go to page ... dialog in full screen view and pressing page up or page down and a filter is set, but no more page is available, a message box is shown. It tells that filtered scroll is used and you may open Go to page ... dialog to change filter. See Go to ...
	Group mode	Toggles group mode. You select a structure element (e.g. an article) in tree view and then you simply click on those zones (or drag a frame to cover more than one) that should belong to the article as well but are not contained yet. The cursor turns to  . See Group mode for details.
	Show empty container	In tree view you may have some nodes like "Illustrations" which don't have any contents because there is no illustration. You may hide those empty containers to make the number of displayed items smaller. Or you want to see them to be able to drag for example an illustration from another chapter's illustration container to the empty one.

3.2.5 Metadata view

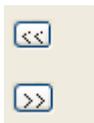
Displays the metadata view. docWizz writes all metadata of the document, its issues, chapters, contributions, illustrations and tables to the metadata section. Here, the metadata can be verified, corrected or edited, for example if alternative text for illustrations is of interest.

	Open previous	Jumps directly to the previous item.
	Open next	Jumps directly to the next item.
	Collapse all	Collapses all branches of the tree structure.



Previous / next metadata

Switch to previous or next metadata (e.g. from a chapter to a previous chapter)



Previous/next metadata for similar entity



Table of Contents

3.2.6 Properties view

A lot of information needed just for a short period of time, document info or history/batch information is displayed on a separate tab called "Properties".

Comments

Already existing comments are listed here. Add a new comment or remove existing ones.

Document info

Shows properties of the document as set in prepare import task.

History/Batch info

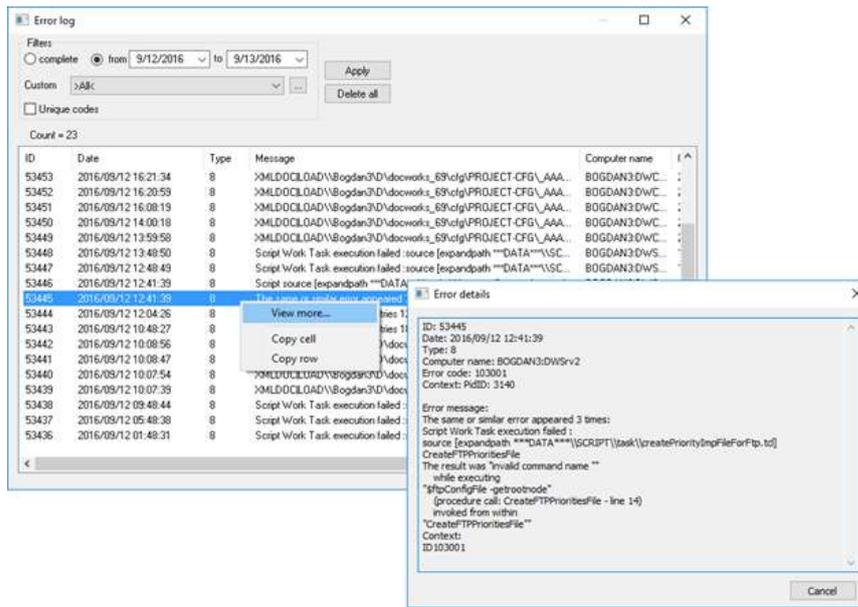
The history shows what's happened before with the document for the selected document (when which job, from whom, where, how long). Values are read from the BatchResults table.

Date	Duration	Task	Result	User	Computer	DW-User	Status	RQA Statu	Error log
25.05.2020 18:13:46	00:03:50	Review structure and text	Discard changes	Oana	ROM-DW-QA-1	dwd	Work	Local	
25.05.2020 18:13:45	00:00:00	Review structure and text	Edited and saved	Oana	ROM-DW-QA-1	dwd	Work	Local	
25.05.2020 18:13:45	00:00:00	Recognize structure and text	OK (step-by-step)	Oana	ROM-DW-QA-1	dwd	Work	Local	
25.05.2020 18:13:37	00:00:07	Recognize structure and text	Edited and saved	Oana	ROM-DW-QA-1	dwd	Work	Local	
25.05.2020 18:13:36	00:00:01	Review issues	Edited and start processing	Oana	ROM-DW-QA-1	dwd	Work	Local	
25.05.2020 18:09:28	00:03:30	Review issues	Edited and saved	Oana	ROM-DW-QA-1	dwd	Work	Local	
25.05.2020 18:09:28	00:00:00	Review issues	Edited and saved	Oana	ROM-DW-QA-1	dwd	Work	Local	
25.05.2020 18:09:27	00:00:00	Recognize issues	OK (step-by-step)	Oana	ROM-DW-QA-1	dwd	Work	Local	
25.05.2020 18:09:23	00:00:03	Recognize issues	Edited and saved	Oana	ROM-DW-QA-1	dwd	Work	Local	
25.05.2020 18:09:22	00:00:00	Review page sequence	Edited and start processing	Oana	ROM-DW-QA-1	dwd	Work	Local	
25.05.2020 18:09:00	00:00:22	Review page sequence	Edited and saved	Oana	ROM-DW-QA-1	dwd	Work	Local	
25.05.2020 18:08:59	00:00:00	Review page sequence	Edited and saved	Oana	ROM-DW-QA-1	dwd	Work	Local	
25.05.2020 18:08:59	00:00:00	Recognize page sequence	OK (step-by-step)	Oana	ROM-DW-QA-1	dwd	Work	Local	
25.05.2020 18:07:46	00:01:12	Recognize page sequence	Edited and saved	Oana	ROM-DW-QA-1	dwd	Work	Local	
25.05.2020 18:07:43	00:00:02	Review zoning	Edited and start processing	Oana	ROM-DW-QA-1	dwd	Work	Local	

Sorting in History Dialog shows the latest entry at top.

Dialog has an **Error Log** button to show all errors related to this document. Column "pages" helps to analyze changes of document content over the processing flow.

The context menu (available on right click) contains, besides the "Copy cell" and "Copy row" options also a "View more" option. This action will open a new dialog with the details of the selected comment / error.



3.2.7 Clip view

Special feature for clipping articles.

- ↶

Undo

Reverses the last operation that you have performed.
- 1:1

Original

The Original function pastes up an article on the target page with the same layout as in the original publication. Automatic reduction can be configured. You can also choose whether the zones on all pages are to be reduced by the same amount or if each page is to be completely filled.
- ☰

Multi Column Block

You can use the Multi-column Block function to paste up the article as a multi-column block across the entire available width of the target page.
- ☰

Match headline

With the Match Headline function paste-up follows the headline structure. This means if the headline is one column wide, the article will have a single column; articles with two column headlines are two columns wide, and so on.
- ☰

Compressed

The Compressed function pastes pages in a compressed mode to save space. If necessary, large images are reduced in size.
- ┌

Column by Column

You can use the Column-by-Column function to paste up zones on the target page column by column so that the space between the header and the footer is filled in the best possible way.
- 📄

Add page

Add an empty page to the page bar.

	Edit	You use the Edit function to edit target pages. Select the function by clicking the Edit button. You can enlarge or reduce the view of a scanned page image.
	Magnetism	Using the Magnetism function aligns new zones placed on the target page as closely as possible to the page margins or previously pasted zones, as if the zones were magnetic.
	Zoom in	Enlarges the image by increments. The (+) key on the numeric keypad performs the same function. Zoom-In is possible to pixel size (one image pixel is one screen pixel).
	Zoom out	Reduces the image by increments. The (-) key on the numeric keypad performs the same function. Zoom-Out is possible to minimum 32 screen pixel width and height.
	Lock zoom	If a zoom factor is selected for one page and icon is pressed, all other pages are shown in the same zoom level (enlarged or reduced).
	Display Entire Page	Resets the zoom and shows again the entire image page. The image is sized to the dimensions of the working window in such a way that its greatest possible spreading "usually vertically" fits into the working window. Shortcut: (Ctrl+P)
	Zoom by right mouse key	Use right mouse button and drag a rectangle inside the page view to zoom once. To zoom to next levels use (Ctrl) key and right mouse button and drag again a rectangle around the desired area. Cancel zoom by the  Display Entire Page button.
	Single page	Displays only one page of the document in the working window. Shortcut: (Ctrl+1)
	Adjacent page	Displays two adjacent pages of the document in the working window – similar to an opened book. Shortcut: (Ctrl+2)
	Display two rows of pages	Displays two rows of pages of the document in the working window at the same time. Shortcut: (Ctrl+3)
	Display multiple pages	Displays multiple pages of the document in the working window at the same time. Shortcut: (Ctrl+4)
	All pages	Displays all pages of the document in the working window. To achieve this, the pages have to be minimized considerably. Maximum amount of pages shown is 50 + x (x means the complete last row is filled with preview pages, even if there are then more than 50) Shortcut: (Ctrl+5)

3.2.8 Custom view

For special customer's requirements a custom view can be setup. Please ask the docWizz support team for individual configurations.

3.3 Rescan

In the Rescan task bad scanned pages can be replaced or missing pages can be added.

In the Rescan task, you will see on the left side the document images and on the right side the rescan control. The control shows at first the task where the document will return to and the list of pages flagged for rescan. Select one of the pages and you will see the detailed description.

You shall now scan all required pages at once in one folder. When finished, click on Load Files and select the rescanned files. They will be shown in the list box.

There are some tasks where images can be checked:

- Prepare Cropping
- Review Cropping
- Review Zoning
- Review Page Sequence

In each of those tasks, images can be viewed and flagged for rescan. Whenever images should be replaced or are missing, the document can be routed to the Rescan task.

Whenever pages are modified, the automated tasks Build Page Hierarchy / Build Hierarchy need to be performed once again. Else other problems may occur at any place. As a conclusion, it is no longer allowed to do such modifications starting from Review Structure without losing work done there.

By the Rescan process, documents are sent automatically to RemoteQA (just if RQAClient is set for this document). If a document shall stay on local system, we recommend to do Rescan in processing in the user interface, step-by-step.

When opening a document in Rescan task, the first paged that has a marker for rescan is displayed automatically.

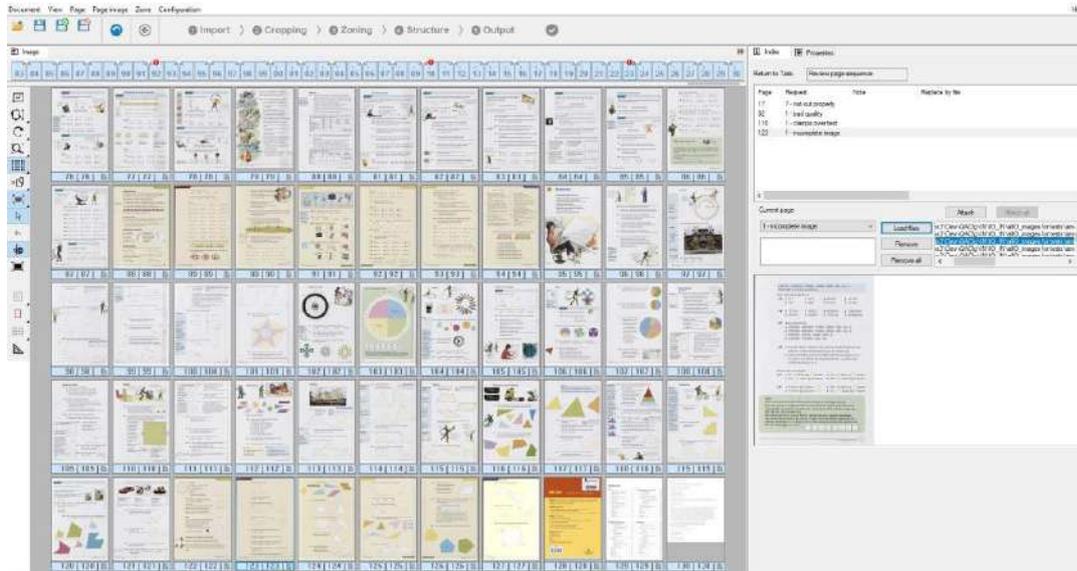
In the Rescan task, you will see on the left side the document images and on the right side the Rescan control.

The control shows at first the task where the document will return to and the list of pages flagged for Rescan. Select one of the pages and you will see the detailed description.

Multiselect is possible on the list control and in case of multiple selection, the rescan_reason dropdown and the edit control are empty.

On selection change or entering a text, the values are written to the selected items in the rescan pages list.

You scan all required pages at once in one folder. When finished, click on Load Files and select the rescanned files. They will be shown in the list box.



Now click on any of the files to see a preview in the window below. Click on **Attach** to assign the rescanned page to the current selected one. The image in the left view will automatically be replaced and page frames will be recomputed automatically.

On **Attach**, in case that there are multiple pages with the same origin as current selection, ask if selected image must be attached to all pages or only to current.

Attach can be done only to pages that don't have an attached image.

A confirmation message in case the selected page already has an attachment will be displayed.

If you have scanned all requested pages at once in the right sequence, you may click on **Attach All** to attach them in one step. "Attach all" is enabled only when the number of the pages that require rescan is equal to the number of images from Rescan list (two pages with the same origin, require a single image for attachment).

Click on **Remove** to empty the file list box. Please note that this action does not delete the files themselves.

Depending on the task from which the document was sent to Rescan task and where the document will return, there will be different processing.

Return to Prepare cropping

When the document returns to Prepare Cropping no additional processing is made.

Return to Review cropping

When returning to Review Cropping the pages are cropped, aligned or split.

Return to Review page sequence

First the images are modified like when returning to Review Cropping task. Afterwards layout analysis and OCR (if selected for the document) is performed. Finally the page sequence will be verified once again.

Handling of 2ups when Rescan

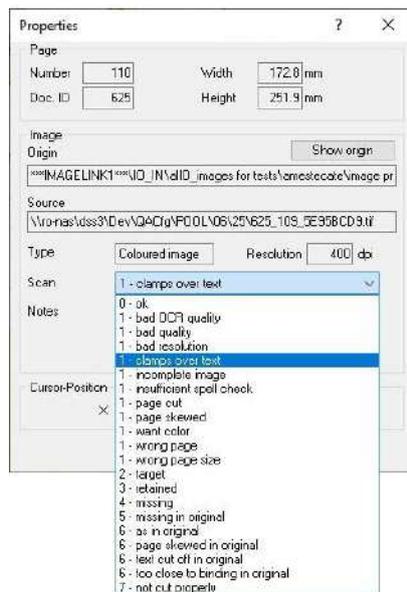
When adding or replacing pages of a book after 2ups splitting has been done, docWizz automatically replaces left and right hand page, if both were flagged for Rescan. In this case you will see the left page with the left frame only and the right page with the right frame.

When finished attaching all files, click on  Process button to provide the document for the operators to continue processing.

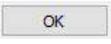
When returning from Rescan, documents will return not to Review structure and text or Review issues but to Review page sequence. The reason is that significant information in the logical structure may be missing.

Flagging images for Rescan

Whenever you detect an image that should be replaced, do a right-click on the image (not on a zone) or do a right-click on the image icon and select Properties.



You will see the page property dialog and may select a reason from Scan drop-down list to tell the scanning operator what the problem is. In addition you may add some Notes that will explain your request more in detail.

Click on  to save your input.

Any page that is flagged for rescan will be shown with a red icon in the page control. So the rescan operator can identify flagged pages easily.



The rescan status and the additional notes are displayed as tool tip on the page icon. According to the different rescan status numbers, different icons are displayed.



If a page is missing, do a right click on the page icon and select Insert Page.

A blank page will be inserted immediately in front of the page you clicked on. It will automatically be flagged as 'missing'.

If the sequence of pages is wrong, you may easily change it by drag and drop the page icons

When you have finished checking page images and you have selected at least one page for rescan, you should select Rescan as next task and then click on  Process button to route the document to Rescan task.

Rescan status

The mechanism for naming and functionality of rescan status has been redesigned. Now the rescan status is a combination of a number (identifying the functionality) and a string which is used for user information.

The predefined numbers are:

- 0 - ok
- 1 - Error (needs rescan)
- 2 - Target
- 3 - Retain
- 4 - Missing (request for rescan)
- 5 - Missing in original (can't be added)
- 6 - As is in original (bad quality, damaged or whatever, can't provide better quality)

Any of those numbers can be used with different explanations, e.g.:

- 0 - ok
- 1 - bad OCR quality
- 1 - bad resolution
- 1 - bad quality
- 1 - incomplete image
- 1 - wrong page
- 1 - want color
- 2 - target
- 3 - retained
- 4 - missing
- 5 - missing in original
- 6 - Original is broken
- 6 - Original is bad print

Displaying Rescan Status

The rescan status and the additional notes are displayed as tool tip on the page icon. According to the different rescan status numbers, different icons are displayed.



3.4 Explanation of workflow steps

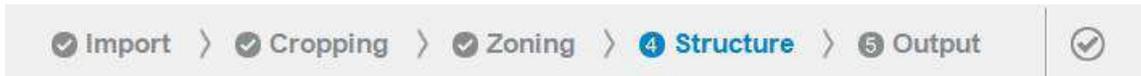


Figure 1: Workflow steps

3.4.1 Process documents

This chapter describes the processing steps in a typical docWizz environment which may not match all working environments.

Processing Steps and Tasks could be:

Step	1 Import	2 Cropping	3 Zoning	4 Structure	5 Output
Task	Setup Import	Prepare Cropping	Review Zoning	Review Issues	Review Output
Task	Review Import	Review Cropping	Review Pages Sequence	Review Structure	

The steps and tasks have three states:

Status	Color	Example
Completed	gray with a hook in gray	
Current	blue with a dot in blue	
Not started yet	gray	

To process the document to the next step you have 3 choices:

- Use the Process button
- or click on the next step
- or double click on the desired step if you want to process a document to a different step than what will be the next task by default.

Before starting the processing task, you will be prompted with a dialog box, where you can choose the Processing mode and enter a comment (optional).

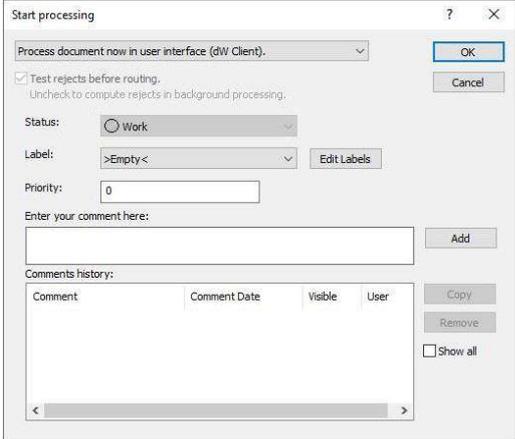
When adding a comment, the date and time is also added for better tracking of the documents. This feature is not compatible with docWizz 6.7. In case a document has comments added from 6.7, the "Comment date" field will have displayed "unknown date".

The system prompts you, if necessary working tasks have not yet been performed or if expected settings are missing or if there are rejects.

The Process button is disabled, if no document is loaded.

The same process is used to route back documents (e.g. to route a document from Zoning to Prepare cropping, select Prepare cropping from the Cropping step).

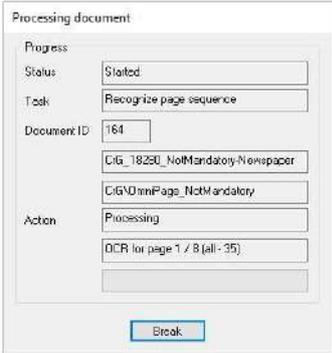
Process documents now in user interface (dW Client)



Remove is used to remove a comment from the current list. Comments cannot be lost - check **Show all** to make all comments visible again. Previously removed comments will be shown in gray.

When adding a comment, the date and time is also added for better tracking of the documents.

If "Process document now in user interface" was selected then you should now see a progress information window displayed.



By clicking on the **Break** button you can stop the process, which causes rejection of the current task. The results of all-preceding processing will be untouched and saved. The document that was stopped in processing will be returned back to the state that it showed before it was opened for the current workflow. For technical reasons the Break operation can take some time to be executed.

Send document to background processing (dW Services)

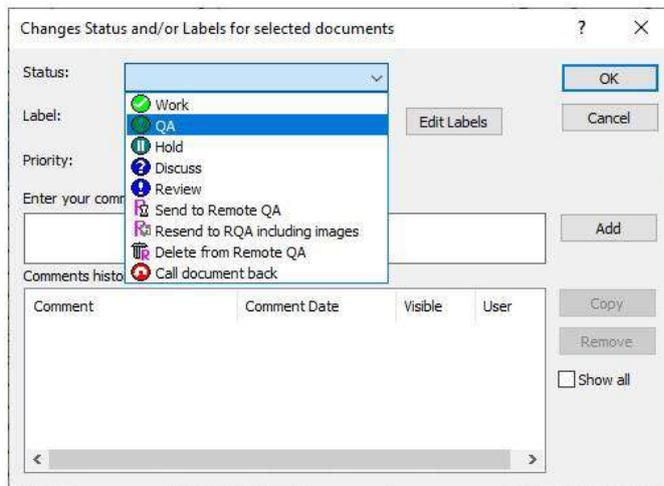
If you select "send document to background processing", you will not see the processing dialog but the document pool open up instead, showing the next documents in the current task. The task sent to background processing will be added to a task queue the processing servers monitor and pick up as soon as possible.

The background processing mode has been configured in order to improve the efficiency of docWizz. Usually OCR is the most time consuming step within docWizz. However, no manual interaction is needed in this process, because of this it makes no sense to wait in front of a computer while it processes this task. Background processing enables you to continue working on other documents and queue them in one or in different stages of processing.

When leaving, for example after work or at lunch, the background processing can be started which means the queued documents are processed automatically. Returning the next morning you will find the queued documents processed up to the task you indicated before.

Change status

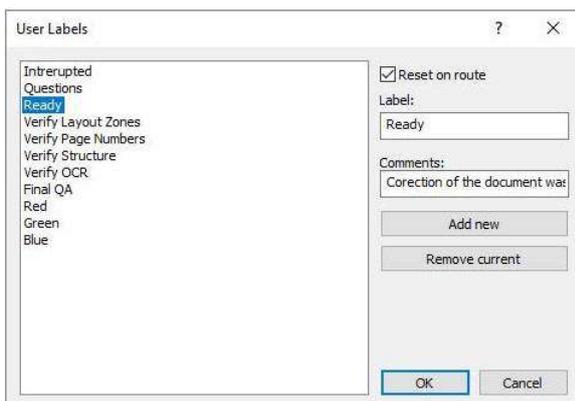
Use the save and close tool  to set status, label and set priority of the document.



Second method is to go to Document Pool and select one or multiple documents. Click on  and the Label dialog will open from there.

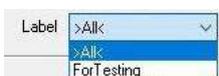
Create a new label or select an existing one and click OK. To remove the label you select the >Empty< value and press the OK button.

Use the  button to add new labels or remove existing ones.



If "Reset on route" is checked, the label of the document will be removed before routing the document to another step.

In the Document Pool you can filter by label. Only documents with the selected label will then be shown in the list.



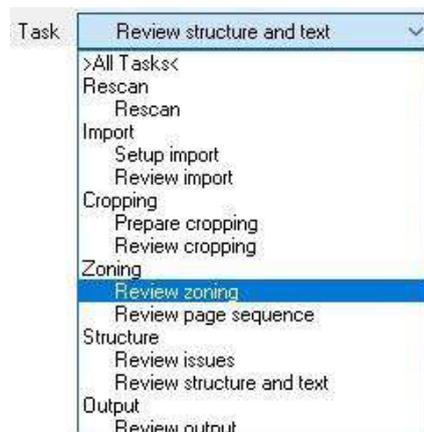
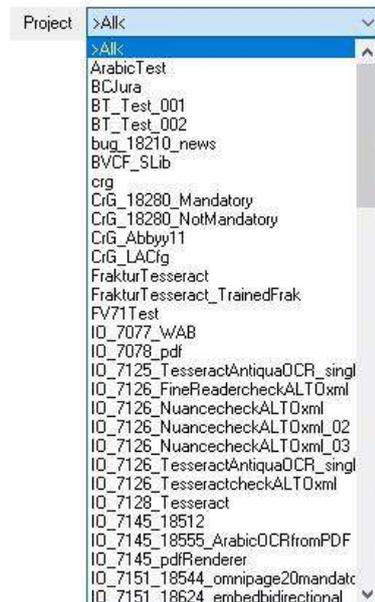
3.4.1.1 Document pool

The document pool shows intermediate results of documents in any task.

Each document is listed along with its unique ID, next task, Date of last modification, Type (serial or monograph) and Title of the document. A lock icon indicates a document currently in use. See **Document pool** for details..

The document pool shows intermediate results of documents in any task. It can be resized to see more data at once.

In order to give a better overview you can apply filters to show documents in the pool. One, two, three or none of the filters can be applied. Use **Task** to select manual QA task.



Switch filter: Just those filters are reset where the requested document does not match the filter selection (or those like custom where it is too complex to evaluate if it matches or not). So the task name is reset as soon the required document is in a different task.

Administrators will see also the intermediate processing steps listed here.



You browse for documents within the document pool by typing in the **document ID**.

The interface allows to define custom filters (administration permission required).

Buttons like **Status/Label**, **Route** or **Backup** are placed on the right hand side of the pool window.

Sort entries by clicking on document's list header.

Tooltip displays information about the document. Use context menu (right mouse) e.g. to copy tooltip text, copy a cell or a row or search for a special string.

ID	Status	Task	Date
59	✔	Prepare cropping	21.10.2013 16:53:00
32	✔	Prepare cropping	14.10.2013 16:49:00
85	✔	Prepare cropping	
33	✔	Prepare cropping	
75	✔	Prepare cropping	
65	✔	Prepare cropping	
64	✔	Prepare cropping	
89	✔	Prepare cropping	
49	✔	Prepare cropping	
46	✔	Prepare cropping	
30	✔	Prepare cropping	

DocID:	59
Task name:	Prepare cropping
Task status:	Work
RQA client:	no RemoteQA doc
RemDocID:	N/A
NextStopTask:	Review page sequence
Workflow:	Image Cut and Analysis
DocumentType:	Newspaper
DocumentTitle:	Tooltip_Test-SK
Project:	Newspaper
Subtask:	N/A
Config:	N/A
Import path:	Y:\CCS Testcenter\4444
Priority:	0
Comment:	N/A
Error/Reject:	Newspaper

RQAClient can be opened as filter. Remote QA documents have also preview in the Document Pool. It is not counted the number of hours, but just the number of days a document is still in RQA, no matter which daytime.

Display only documents from the last ... days. The amount of days is configurable. It is recommended to set a limit here, to reduce database loading time.

Columns

There are columns available in the dW pool window. Please find below the columns and their meaning.

ID	Status	Task	Date	Type	Pages	Priority	Group	T
8	✔	Review zoning	11/15/2019 10:32:00	Seven One Newspaper-Newspaper	8		Si	
9	✔	Review cropping	10/23/2019 10:05:00	Seven One Monograph-Monograph	36		M	
11	✔	Review page sequence	11/19/2019 12:16:00	Seven One Simple Monograph-Simple Monograph	36		Si	
12	✔	Review structure and text	5/26/2019 1:57:00	Seven One Simple Newspaper-Simple Newspaper	28		Si	
13	✔	Review output	10/24/2019 3:52:00	IO Mono7_1 single-Monograph	15		IC	
15	✔	Review output	11/09/2019 2:51:00	Seven One Monograph-Monograph	10		Si	

Workflow

Shows the current workflow, how the document should be processed.

StopJob

Indicates the next task where the documents stops and where an operator needs to do the verify task.

ShareIndex

Index of the POOL the documents are stored in. Of importance especially when technical issues are known for one share and due to this, related documents should be focused on or ignored (one pool share is running out of space / share to be cleaned / cannot be accessed).

RQA Client

Indicates which remote QA location is working on the Review tasks.

RemotelD

Indicates the ID used on the remote QA location.

Config

The CONFIG and PROJECT can be defined individually. So ProjectName and ConfigName can be different. By this multiple projects can use same project configuration. In case this is used, it is confusing for support and clients as to why documents behave different than others. Now it can be shown in the Pool view which configuration belongs to a project.

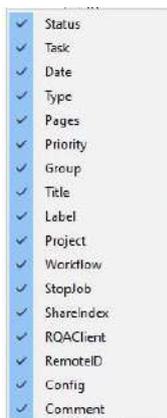
Comment

Indicates the comments which are made during processing either via "Status/Labels" icon or when the document is open in dW via the comment field in the index view.

The "Note" is named and linked to the "Comment" field existing which is right now ONLY in the ID.xml / tool tip. The search for the comment is adapted as well so the database is used instead of the ID.xml files for speed improvement.

Sort columns

Click with right mouse on a column headline to open the list of all available columns and mark the ones you are interested in.

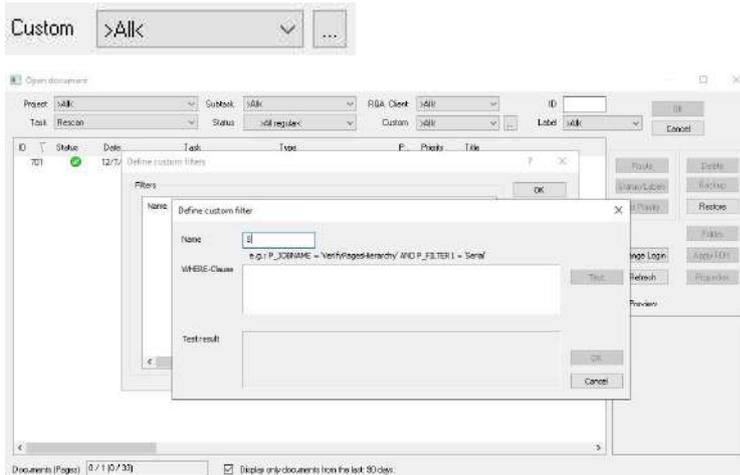


Sort the column order using drag and drop to move the list headlines to the desired place:

Custom filters

To have a more specific pool view, custom filters for the Pool dialog are configurable. They contain a pair of displayed name and a fragment of a "where" expression from an SQL select statement. Administrator users can define new filters within UI. Filters will be selectable in a combo box and can be combined with any other filter.

Log in as administrator. You will see a button with three dots to open a separate window to define custom filters.



Each document is listed along with its unique **ID**, next **Task**, **Date** of last modification, **Type** (serial or monograph or newspaper) and **Title** of the document. A lock icon indicates a document currently in use.

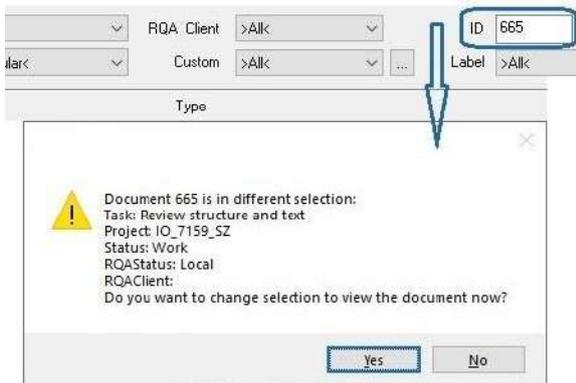
Whenever a task has been sent to the processing queue, the next task is an automatic process. All these are starting with **Detect...** (exception: SplitDbIPages) or **Build...**, so the operator can identify prepared tasks easily.

If an entry starts with **Verify...** or just **Scan** the related document is not prepared for batch processing but instead waits for an operator to be opened up.

The pool folder structure can be extended to two levels of folders to improve performance on mass digitization projects.

When changing a filter manually it is checked whether new document type is available.

ID



Enter ID and press Return. You will get an information window that shows details of the document (Task, Project, Status, RQAStatus, RQAClient). Task and custom is missing here, because Custom Filter is not a document property. A document may have multiple active tasks at a time.

Search by ID sets filters for Task and Status: This is done to reduce the time needed to display the documents, especially for big environments. When a document is searched by ID, the Task and Status of the document are set as filters, reducing the amount of documents needed to be displayed.

Open Document

To open a document, select it with a mouse click and press the button or simply double click on it.

When entering a document and the document is currently not visible, a message box shows where to find the document. Click yes to have the pool filtered to show the document searched for.

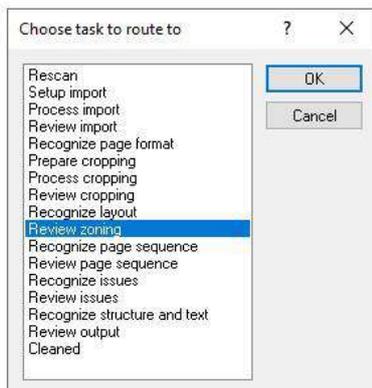
Double click on task name

Opens pool view with according task filter.

When monitoring the pool you sometimes have to check the reason for a number of documents in a task. A double click fastens the actions on the task in the root node view (task names, number of documents, priorities) the pool view will be opened with the filter set to the selected task. If you manually open the pool view you have to first wait for the loading view for "All" documents which can take quite some time depending on the amount of documents in the pool.

Route Document(s)

To route one or many documents, select the documents (note: multiple selection requires documents to be in the same task!) and click the  button. Then, choose the task the document(s) should be routed to. Routing might become necessary if a certain correction has been forgotten or a change in correction policy took place.



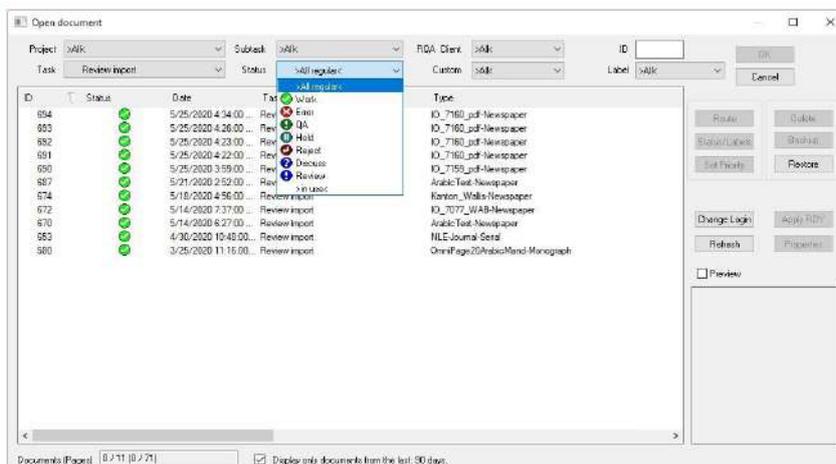
Routing of documents in Free Pool Data status is not allowed.

Whenever a document is routed back (e.g. Review structure and text -> Review page sequence) you can do some changes, process to Review issues. As soon as processing starts, all layout analysis results from further tasks are deleted and no longer available unless recreated.

To route documents one step back from outside the document pool you can also use the "Back" button in the top bar.



Status



>All regular< - displays the documents from all regular statuses (Work, Error, Critical error, QA, Hold, Discuss, Review).

>in use< - shows documents currently in use.

As admin you see additionally:

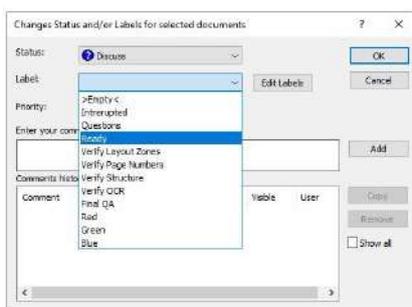
>All cleanup< - displays the documents from all cleanup statuses (Free Pool Data, Reduced Pool Data, Restore Pool Data).

>All RQA< - displays the documents from all remote statuses: On Manager: Send to Remote QA, Delete from Remote QA, Resend to RQA including images, In Remote QA, Remote QA done, Prepared to be sent back, RQA done, sent back, Call document back. On Loader: Send to Master for review.

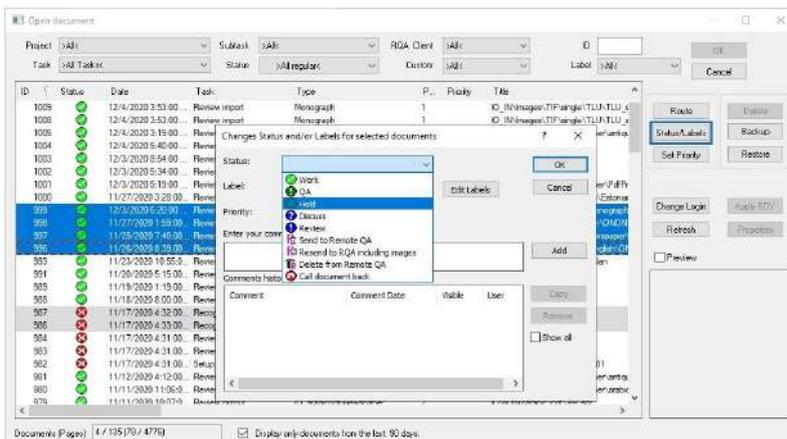
Change Status

Select one or multiple documents from the list. Click then on **Status/Labels** and the Label dialog opens from there.

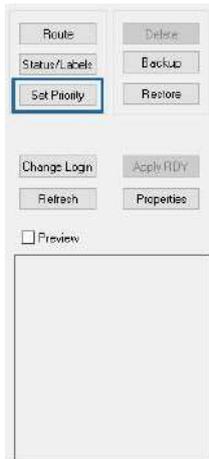
Create a new label or select an existing one and click OK. To remove the label you select the >Empty< value and press the OK button.



After selecting one or more documents in the Document Pool, status can be changed by clicking the **Status/Labels** button. You can also enter a reason or comment. A progress bar is shown when changing current selected documents.



Set priority



The **Set Priority** button enables documents to be processed with higher priority.



Change login

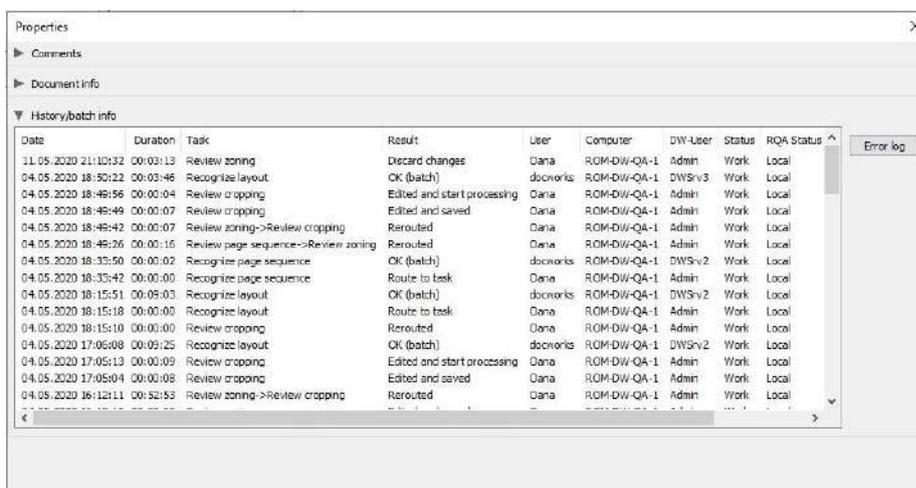
Change login by **Change Login** to another user (or admin) e.g. with permission to delete documents and switch directly to the project's IN path.

Apply RDY

In regular cases, docWizz knows, which rdy template has been used. Use the **Apply RDY** button when performing import via dedicated settings inside dWScanClient or other scripts. This feature is restricted to administrators. Find more information in the docWizz reference manual.

Properties

The **Properties** button shows what's happened before with the document for the selected document (when which task, from whom, where, how long). Values are read from the BatchResults table.



Sorting in History/Batch info part shows the latest entry at top.

Dialog has an Error Log button to show all errors related to this document.

Column "pages" help to analyze changes of document content over the processing flow.

Refresh

The  button updates the document list, so operators can indicate what documents have been processed recently and are ready to work on. The preview check box shows the first page of the selected document.

If update document status in pool database fails, retry is performed.

Delete Document

To erase a document, select it and press the  button. This feature is restricted to administrators.

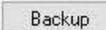
Restore Document

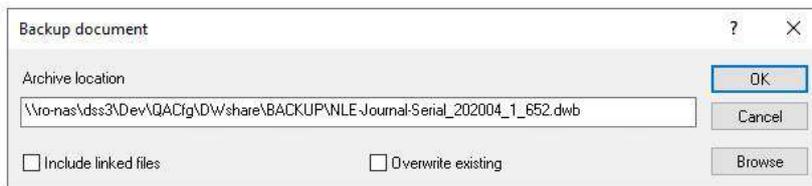
To restore a document from a backup file, push the  button. Go to the directory where the backup files can be found and double click on the backup you want to restore. Choose whether you want to perform the restore immediately on the workstation or if you want the processing servers to perform the restore task.

Note: This feature is restricted to Administrators.

If you add a couple of backups to the restore queue and want to view this queue, you can do so by using the (Alt) menu 'Configuration' - 'Maintenance'. Here, go to 'Backups to restore', you will be able to remove tasks here.

Backup

To create a backup manually, hit the  button. Multiple selections are possible. Choose the destination where to save the backup and decide whether you want to include linked files or not. This means, whether source images are included or not.



Clicking on the Open button will open the pool, import, or project configuration folder either in Total Commander (if found in "program files" folder) or Windows Explorer.

Folder

For administrators it is often needed to have a look into a documents pool folder.



Clicking on the Open button it will open the pool, import or project configuration folder either in Total Commander (if found in "program files" folder) or Windows Explorer.

Note: This feature is restricted to administrators. Button is only visible if you are logged in as administrator.

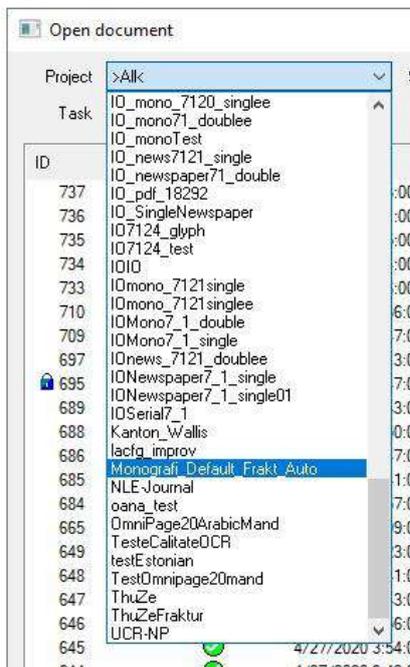
Preview

Mark the preview check box to show thumbnail of the document.



Select Project

If you select one project, only documents belonging to this project are shown. Select >All< to see any document of any project.



Locked documents

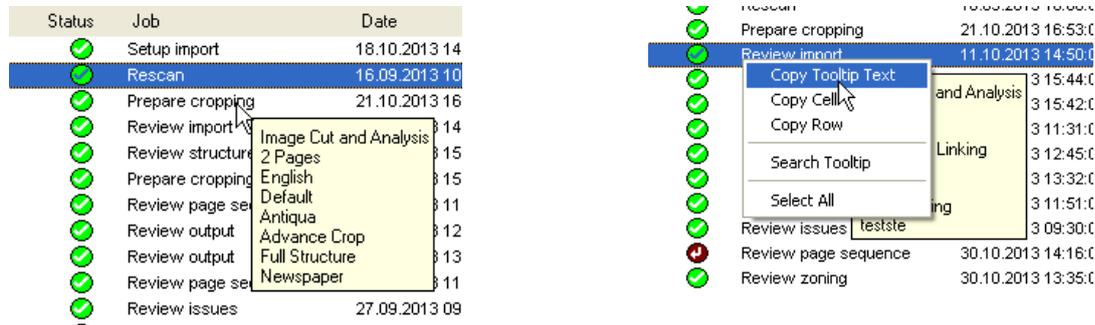
Locked documents are shown with small blue lock icon  at left hand side. This means, that the document is currently used by another user (or by you).

When a document is in use, it will also show which computer the document is in use by (click with right mouse button to show).

Note: According to your entry in the Current task window, the Open Document dialog (you can also open with the Open... function in the Document menu), only shows the documents in the pool that fit to your settings and are in a state of processing according to your current choice.

Context menu

A context menu is available to copy cells or rows. It is also available for multiple selections of documents in the pool view.



Copy Tooltip Text copies the text from the tooltip into clipboard and can be inserted into another application (e.g. Editor, Notepad).

Copy Cell/Copy Row copies data of all selected documents from the column/row where a right-click has been made.

Search Tooltip opens a dialog to enter a search string. All documents that contain the string in tool tip or title are selected (just from current filters). This operation may take a while because it needs to read each id.xml file. You may abort the function by pressing the (ESC) key.

Select All selects all documents at once.

Close / Open

After closing and reopening the pool dialog, task selection box from the document pool dialog is set to the current selected task from the workflow bar. It will not keep the last task filter selection or status filter.

Display only documents from the last 40 days

If checked: Displays only documents from the last x days; Unchecked: Displays all documents.

Search by user

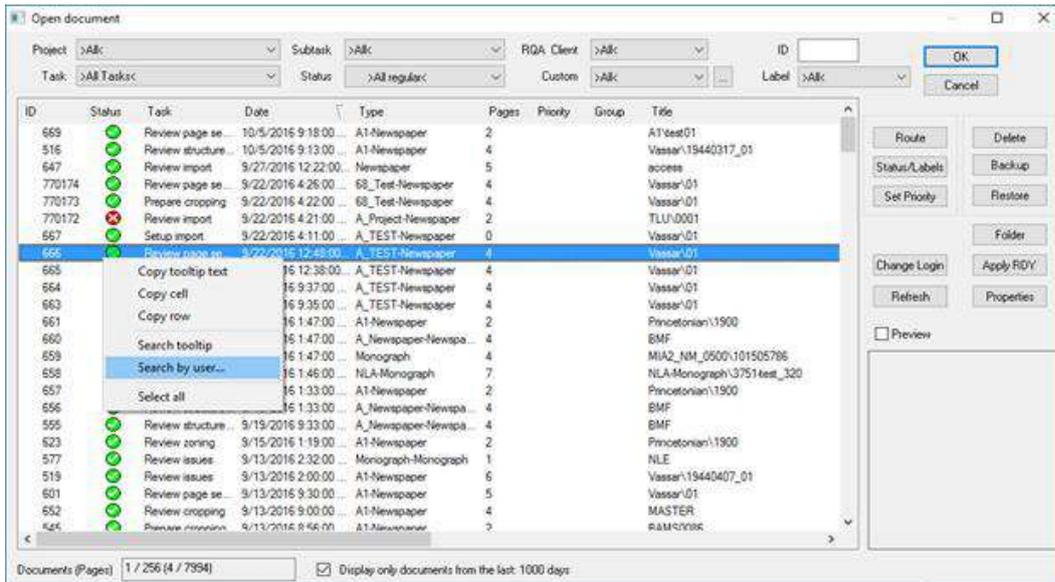
The action is available in contextual menu (right click) and opens up a dialog where you can select:

- User type (docWizz or Windows)
- User name
- The time interval to search in
- Possibility to show only "relevant documents":

The user has made at least one action on document.

Specify the minimum time spent by the user on document.

The search is done only among the documents that are displayed in Document Pool. The ones that are found to match the criteria will be selected.

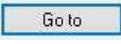
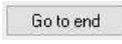


3.4.1.2 Go to ...

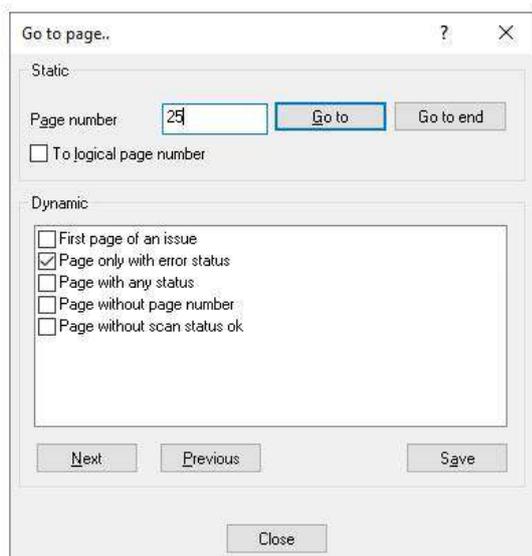
Helps you move from one open page to another within the open page batch. This function can also be accessed from the page bar where the open pages are displayed. Shortcut (Ctrl+G)

In **Image view** on left hand working window go to the **Page menu** and select **Go to page** or use the (Ctrl+G) keys to go to a certain page number or to a page that fulfills a selectable, dynamic status.

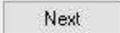
Using the **Go to page...** dialog in full screen view and pressing **page up** or **page down** while a filter is set, will display a message box if no more pages are available. It will say that filtered scroll is used and you may open the **Go to page...** dialog to change the filter.

Type the page number and click  to go to the certain page number and click  if you want to see the last page.

Use of **To logical page number** is possible, too.



The static **Go to page number** is available in all interactive tasks.

In the **Dynamic** area the dialog shows options depending on the current task and the document type (newspaper, monograph or others). You can select one or more option there and use  or  button to switch to a page that has the selected status.

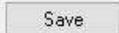
Examples for dynamic page search:

Available for newspapers and serials but not for monographs:

- First page of an issue.

Available for newspapers, serials and monographs:

- Page only with error status.
- Page without page number.
- Page without Scan status "ok".

The  button allows you to save modified filters without executing one of the go to commands.

3.4.1.3 Merge (Stitching)



Merge with next page "Stitching"

This is a special function used for "stitching" – for the semi-automated creation of complete page images. It is available when configured. Stitching is used to stitch two half scanned pages into a single one. Used for special customer projects.

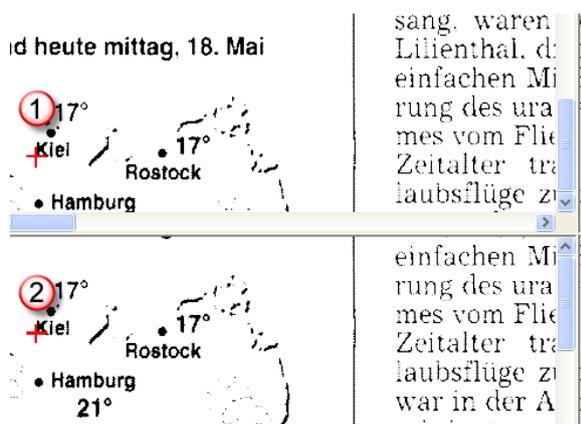
This feature is available in the "Review import" task.

Sometimes pages are not scanned or microfilmed as a single image. Instead two overlapping page images are created. Those images might differ in zoom and skewing. As well the pages might be curved.

A special algorithm handles the overlapping area by fading the brightness. This algorithm is available for grayscale images only.

Select **merge with next page** tool . The mouse pointer changes automatically into a hair cross icon that you can place accurately with one click. Renewed clicking shifts the red hair cross to another place and deletes the past measuring point.

Click on two significant edges on each of the partial images. Then the system automatically creates the final image.



You can move the red cross using arrow keys to be more precise. The arrow works on the selected window.

For a more exact positioning of the measuring points you can increase the preview magnification using over the magnify tools or or change to the whole page size view. Set the measuring points in each case as exactly as possible in the same place of the overlapping range in the upper left and lower half (1, 2).

Confirm the measuring points with , the arrangement window is closed and the joined page is opened in docWizz. If you click , the merge window is closed, without storing the merged and joined page.

The stitching "stick oversize" toolbar contains the active items as listed below:

Image OK

If the quality is satisfactory, select **Image OK**. Then the final image is created, and the stitching window is closed. docWizz takes into account to merge only white spaces between characters to avoid damaged characters whenever possible.

Cancel

The stitching window will be closed without saving any image.

Zoom in/out

Zoom in enlarges the image, zoom out shrinks the image view.

Display entire page

Shows the entire page. Shortcut: (Ctrl+F)

Zoom 1:1

Zooms the image in 1:1 format. 1 pixel of the image is represented by 1 pixel of your screen.

Display color/grayscale images

Toggles between the display of images in grayscale/color or black-and-white. Usually, page images are displayed in docWizz as black-and-white (1 Bit) images.

3.4.1.4 Knife - Polygon



Polygon zone

You can use the Polygon zone button, or press the (Ctrl+P) key, to create a polygonal (many-sided) zone for clipping a text column with a formed setting. Once you have activated the function, place the mouse pointer at the point you want the polygon to begin and click the left mouse button to indicate the starting point of the polygon. When you move the mouse a line appears on the screen beginning at the polygon's starting point, and behaves as if it were magnetically attracted to the mouse pointer, it will follow it, but without leaving the north-south and east-west orientation. As soon as you press the (Shift) key the line moves at any angle you want around the starting point and connects it directly with the mouse pointer.

Each click of the left mouse button adds a new point to the polygon.

When you release the (Shift) key the line currently being processed becomes horizontal or vertical starting from the last fixed corner; pressing the (Shift) key again allows you once again to adjust the angle.

Use the Backspace key to remove the fixed points in reverse order, thus deleting parts of the polygon outline again. Press the (Esc) key to delete the entire polygon.

To close the polygon, press and hold down the (Ctrl) key and click the left mouse button. Alternatively, you can press the (P) key after setting the last point; this then connects this and the starting point thus closing the polygon.

The polygonal zone is given a yellow background and the type is automatically designated Paragraph. If necessary, you can change the type by placing the mouse pointer in the zone and pressing the right mouse button. You then make the necessary changes in the Zone window.

You can add additional points to any polygon. Place the mouse pointer on the edge of the zone where you want to add the point. As soon as the pointer appears as a double arrow press the right mouse button. A new point is now added where the mouse pointer was positioned.

If you want to delete a point from a polygon, place the mouse pointer on the point you want to delete, and, when it appears as a four-headed arrow, press the right mouse button. The point is removed and the edges are redrawn.



You can also delete segments of a polygon without deleting all the points that lie in them individually. To do this, place the mouse pointer on the starting point, and when it appears as a four-headed arrow, press the (Ctrl) key and click with the left mouse button. Move the pointer to the end of the segment you wish to delete and click on this point or on any point on the edge and click again when the four-headed arrow  appears again. You can see the line joining the two points that you have just clicked on. All the points in between are deleted and the two points are joined with a straight line.



If you want to move a point of a polygon, place the mouse pointer on the point you want to move, and, when it appears as a four-headed arrow , press the left mouse button and move the point to the desired position, holding the button down while you move the point.

To move a polygon edge parallel to its original position, place the mouse pointer on the edge you want to move until it appears as a double arrow  . Holding the left mouse button down, move the edge parallel to increase or decrease the size of the polygon.

To change a polygonal zone into a rectangular zone (or vice versa)



Select the zone and press (F8):



To avoid overlapping of zones press (F8) again.



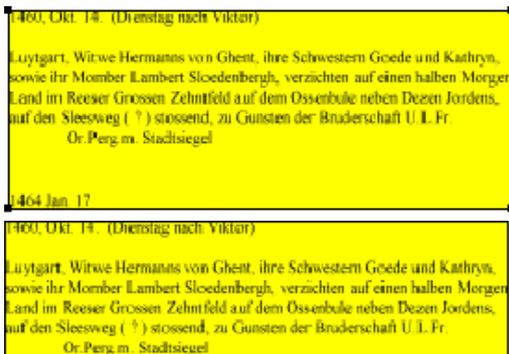
The Knife function is used to split zones. Select the function by clicking the Knife button or the (F9) key. The mouse pointer changes to a knife . Move the knife in the active zone to the position where you want to make the cut. Clicking the left mouse button with a slight movement of the mouse cuts the zone into two pieces. Exit the function by pressing the (K) or (Esc) key.

A zone cut here is a normal cut that cuts only when an empty area is found.

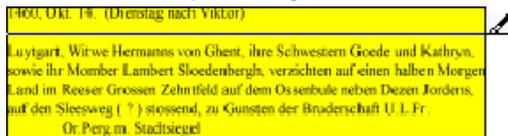
The Knife function has the following options, the mouse cursor changes accordingly.

 Horizontal Cut

Move the knife to the position on the zone edge where you want to make the horizontal cut. Click the left mouse button. dW tries to make the cut in the nearest white space.



If you hold down the left mouse button, you can see a horizontal line that you can move up and down. You can move the line to the exact position you want and when you release the mouse button, the cut is made at that point, regardless of the zone content.



↖ Vertical Cut

Holding down the (Shift) key changes the mouse's appearance to  and the knife appears in a vertical position. Move the knife to the position on the zone edge where you want to make the vertical cut.

↙ Make polygonal cuts

You can also make diagonal cuts, or cut irregular shapes like polygons. Position the mouse pointer on the edge of a zone. The mouse changes its shape and a small staircase appears beneath the knife.  Click the left mouse button to position the first point. Now move the knife to the other positions where further points of the polygon are to be placed. Each additional click sets a new point. By pressing the (Shift) key at the same time you can move the mouse vertically. By pressing the (Ctrl) key at the same time, you can move the mouse like a stair in 90 degree angles. When all the points are positioned to your satisfaction, finish the cutting process by clicking on the edge of the zone.

Note: You can drag a small zone on a polygonal and make them rectangular. Rectangular zones look like grayscale. A polygonal shows the image like black and white and only a few pixel are shown.

Button  Image Zones Gray is active for polygons and rectangles.

The idea behind is that having that show color illustration button, overlaps with the transparency of the zone, polygonal illustrations, by definition, are filled with the color, while rectangular illustrations are transparent with a bigger border.

3.4.2 Import

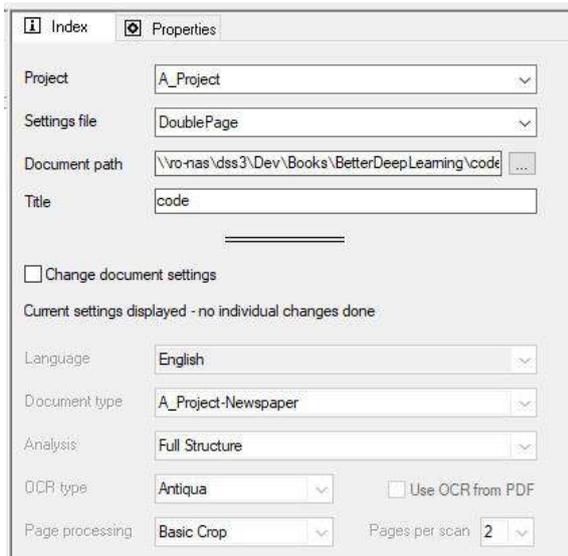
In this step you import documents from network storage. Depending on your environment and your workflow there are different methods to import documents into docWizz.

3.4.2.1 How to use the Setup import task

Previous name of this task was Prepare Import.



The **Document Import Setup** dialog on the right allows you to add descriptive information to a document. The added information can be used for further processing.



What is the difference between Setup import and Review import? Why are both jobs needed?

When document pages are available they usually are displayed in the left working window. You start working with a document either by scanning the printed pages using a scanner or by importing page images that have been scanned before and are available as image files.

Project

Select project and workflow with the appropriate RDY file from the drop-downs.

Document path

Enter or select document path. This path should lead to a server (starting with \\servername\) and not to a local path (like e.g. C:\).

Title

Fill in the working-title of the document. It does not necessarily have to be the original title of the source document. During processing you will always find the document - for example in the document pool - under this title.

You can fill in the title by typing it or using the mouse to draw a zone or drag & drop to the title field. The text within the zone will be read supported by the systems OCR engine. Put the cursor in the source page on the zone that marks the title of the document and press the left mouse button. The cursor changes its shape . Hold down the left button and drag the zone from the source page to the document dialog and release it in the Title window. The string, recognized by the OCR engine, will be filled into the window

after a short processing moment. In case the engine is not able to recognize the text with the necessary confidence, it automatically opens the Correction interface, where you can see and correct the text.

Note: If Backslash "\ " is used in title field (Start step): sub folders will be created according to the use of "\ ". However if "@title" is used also for file naming of images and METS and ALTO "\ " will be ignored (deleted from file naming).

Change document settings

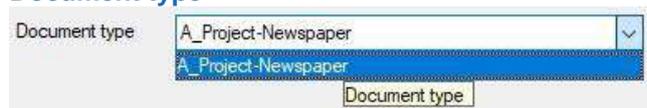
Change document settings

If not checked, the default settings will be used.

Mark here if you want to edit and enter individual settings for the document. The fields and drop-downs will become active.

Switch between checked and unchecked to let the system update your value changes.

Document type



docWizz is prepared to deal with different types of sources, e.g. Monographic, Serial or Newspaper documents.

Note: The list of document types is extendable due to project requirements and may look different on your system.

Language

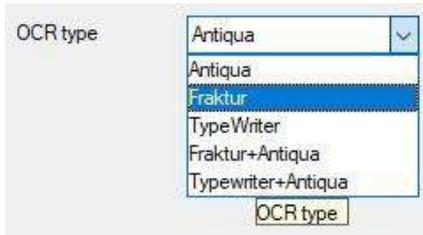


docWizz is prepared to process documents in many different languages. As there are so many, the most used languages are pre-adjusted in system configuration. Select the language of the source using the drop-down-list in the Language field. The Language setting is important for further processing, it influences the character recognition processes (OCR) of docWizz and the use of string lists and dictionaries. If a document contains more than one language it is possible to select here more than one language.

We recommend to use 3 languages maximum.

OCR type

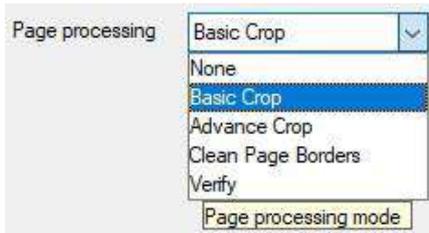
You can change OCR-Type to:



Select the main Font type of the document.

Page processing

Here you select, what the system should do so that pages look better.



None

No page processing will be done

Basic Crop

Black or noisy borders will be removed, the position of the printed text will not change.

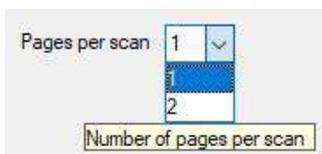
Advanced crop

Margins out of the red frame will be cleaned (filled with white or background color depending on configuration). Image itself will not be changed in size and position (no deskew). There are a variety of tools to edit page frames in Cropping step.

Clean page borders

Black or noisy borders will be removed, the position of the printed text will not change.

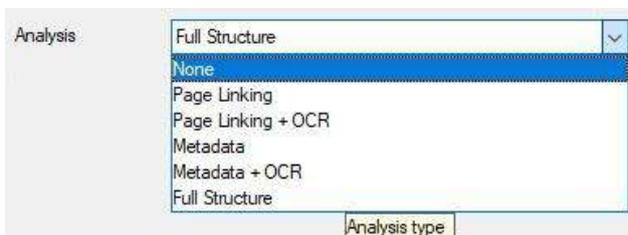
Pages per scan



Select whether one or two pages are scanned at once.

Analysis

You can choose different analyse settings.



None

No OCR or structure analysis of pages will be done

PageLinking

Page numbers will be detected and linked to the images.

Page Linking OCR

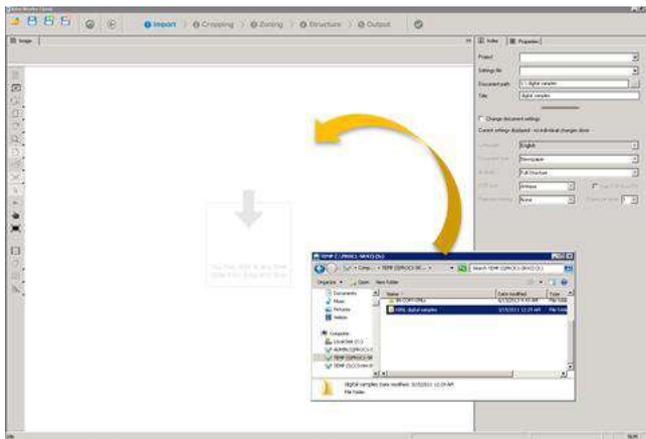
As before, but also OCR will be done.

Full Structure

As before, but also the document structure and hierarchy will be detected.

“drag and drop” functionality

Folders can be dragged in Setup import – this will set the Document path and Title to the path of the selected folder. Dragging image(s) will set as the path of the folder containing the image(s). In case the document already has a path and title, dragging a different folder / image in Setup import will change the path to the path of the new folder.



Setup/Review import – do not allow processing if document is not complete. In case the document is incomplete, when pressing “Process”, “Save”, “Save and close” or another task, a message will be displayed: “The document is not completely configured. Please also fill (empty fields here) or discard changes” with only one button: OK; after pressing OK, nothing happens and you have to enter the missing entries.

Working with 2ups

For correct processing of double pages you have to split them into single pages first. Using the **Scan** job you can load double paged images.

Split Pages

Divides a page image into two parts. This function is only available when you have loaded landscape formatted page images. It is useful when you capture double pages and integrate them into a single page document. In this case the double pages could be split into two single pages in a semi-automatic procedure.



1st Step

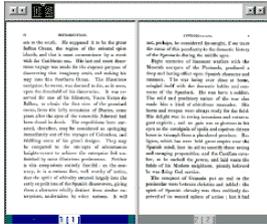
Choose a scanned page and activate the function by pressing the button  , docWizz detects automatically the print spaces of the single pages and presents the result for you to verify.



The red frames indicate the automatically detected areas of the page. You can change the frames using the mouse.

2nd Step

Clicking the button  again saves your changes and divides the scanned page into two document pages.



Unnecessary borders around relevant areas are deleted automatically. Page numbers are assorted automatically.

This method is used when you process a complete document with Page Processing setting 'Double Pages' as well.

Using the  button you can view several pages in the left window and check the whole batch of pages from the beginning to the end. You can manually and individually correct wrongly positioned or sized frames using the mouse to grab and slide the sides of the frames. Depending on where you place the cursor on the horizontal or vertical side of the frame, it changes its appearance to   and allows movements in the direction of the arrows.

You can also correct crookedly scanned pages using one of the deskew  (manual deskew) or  (automated deskew) options.

Once the frames have been checked and corrected you can use one of the



buttons and let the system calculate the maximum size of page dimensions in the current document. You can manually interfere and change the calculated values to those of your choice by typing the desired values into the Width and Height input fields. The system expects values in 1/10 mm in these input fields.

Start processing by pressing the  button.

Note: In production lines, this process might be performed by processing servers and you will not see the progress dialog.

The process of splitting pages is performed automatically. The system shows the progress of processing in the **Processing Document** information window.

According to the page width settings, the system creates new blank pages and the content of the red frames is pasted to these pages.

After **Processing** the number of pages in the page bar has doubled and each page symbol now represents a single page.

3.4.2.2 How to use the Review import task

This task is used to create and configure new documents.

Project

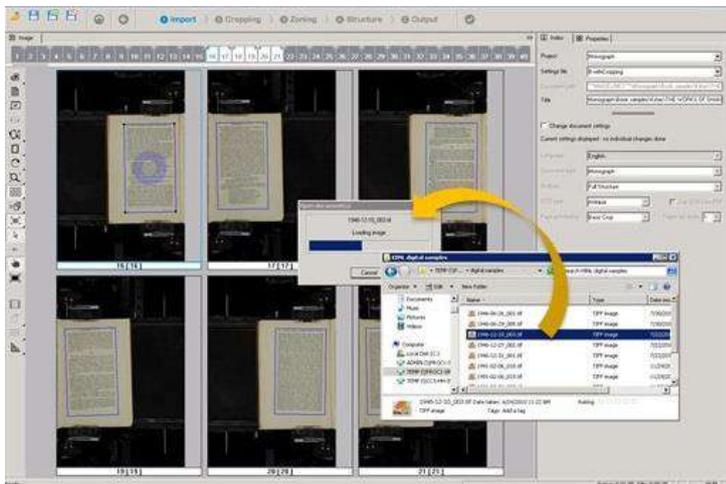
Select the desired Project from the dropdown (left column) and the desired RDY file from the right dropdown.

Press the Open File button and select the desired images you want to add to the document. Individual pages can be added, no need for the IN folder then.

When everything is completed, press the process  button.

"drag and drop" functionality

Files can be dragged in Review import.



Split documents

Large batches of huge images can't be processed as one single document in docWizz, especially when layout analysis and OCR is required. For that reason, those batches need to be split in the Review import task.

Although there is no precise limit for the size of documents, it must not exceed a certain size, because of memory usage and load/save speed.

Note: This feature makes sense for newspaper scans only. If books are split, then several documents will be produced in docWizz. This will also result in an export folder for each document with its own file naming. This is not ideal because ultimately one would have to unite the two or more documents as one again. Thus, normally books are not split.

To split a document follow the instructions below:

Open the batch in **Review import** task. Then scroll to page with the number that is close to the requested limit or use (Ctrl+G). Search for a page where you want to split the document (for newspaper and serial it should always be the page where a new issue starts).

Select the page and click on **Split Document**  in left toolbar or use (Ctrl+T).

You will see all pages starting from here in the page bar having a red cross. This shows which pages will belong to the next part of the batch.



If you clicked on a wrong page, select the right one and choose **Split Document** or use (Ctrl+T) once again. If you want to get rid of the document split setting, select the first page and click here on **Split Document**.

Once you have set the correct split position, click on Process  or press (Ctrl+Shift+P). The first part of the batch is then taken as a separate document and sent to batch processing. Please note that the processing in user interface is not available here and the document will always run in batch, no matter what you have selected.

After clicking on Process, the first pages are removed from the Review import task and you will see just the remaining pages. Repeat this procedure until the number of remaining pages is smaller than the maximum number. For the remaining pages, click on Process and they will become the last document of the batch.

Recommended maximum sizes are:

- 2000 pages for normal books (one, maximum two columns)
- 1000 pages for books with more than two columns
- 500 pages for newspapers with small page sizes
- 50-200 pages for newspapers with big sizes
- If just image processing is done, documents shall not exceed 3000 pages.

Note: The numbers are valid for single pages. If double page splitting is done, choose half the number of images in Import step!

Handling 2ups (double pages)

There is special handling when processing 2 ups. Here the first page that is selected for the next part is duplicated and added to both documents. This is required to get complete issues. In **Prepare Cropping** you need to remove the page frame from the image that is not part of the issue on first and last pages of the document.

Please note that for technical reasons the last part has the lowest document id and is shown as the first in the list. We recommend not to split directly before or after an alternative page.

When you look in the document pool, you will see the partial documents, having the same title with an extension which shows the part number in brackets.

Please note that for technical reasons the last part has the lowest document id and is shown as the first in the list. We recommend not to split direct before or after an alternative page.

Import PDF docs

Use hidden text of digital born PDFs instead of executing OCR:

Direct import for PDF documents is possible. docWizz is able to use hidden text from PDF files. Text inside the PDF might be used instead of OCR. This hidden text exists in born digital PDF's or in some cases it has been added manually to PDF's of scanned images.

By taking existing hidden text into the docWizz workflow OCR is no longer necessary. Thus OCR errors can be avoided and processing is speed up tremendously.

In case some zones do not contain hidden text OCR will be performed as usual. The same is true if the whole document lacks hidden text.

To activate the feature please edit the project configuration (refer to the Project Configuration documentation of docWizz Reference Book). Please perform the following changes:

Refer to the "Processing" element of the <project_name>.rdy file.

Edit (or add) the line <OCR> so that it quotes <OCR SCRIPT="OCR-PDF" />

3.4.3 Cropping

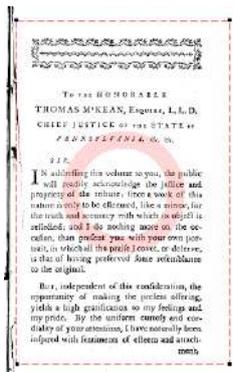
The cropping step is used to crop images, clean borders, split 2ups and finally check if the pages are cropped correct. Cropping can be done as Basic Cropping in an fast and easy way or as Advanced Cropping with more tools to crop pages.

Page frames

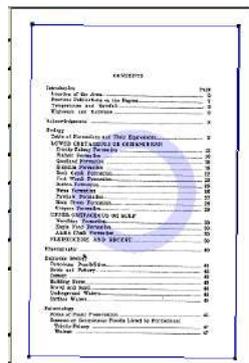
Active and selected pages have circles in the center. The color of the circle is dependent on the frame color (red, blue, orange).

The colors of the frames are configurable. See in the manual [docWizz ReferenceBook](#), chapter [Change colors of frames for Cropping](#) step.

If the frame is from a regular page (red) the circle in the center is red



If the frame is from an individual page (blue) the circle in the center is blue



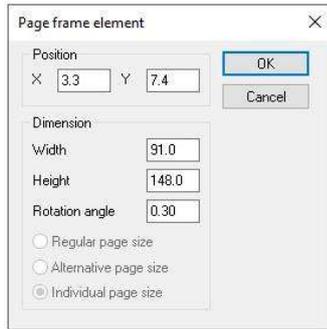
If the frame is from an alternative page (orange) the circle in the center is orange



Frames are selected when:

- Current page always has the frame selected, in case of 2 ups, the left frame is selected.
- An action with the mouse is made on the frame (resize, reposition).
- A right click anywhere inside the frame.
- On 2 up pages - using (Ctrl+R) for right frame and (Ctrl+L) for left frame.

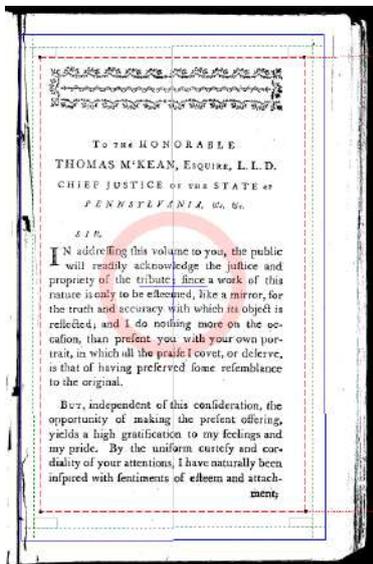
The Page frame element window is accessible by a right clicking on a frame margin.



For 2 up pages, the selection can be easily changed between frames by using (Ctrl+R) for right frame and (Ctrl+L) for left frame.

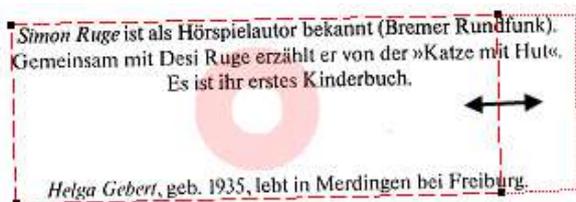
The Content area (green frame) is connected with the final page frame (blue) and is only displayed together with the blue frame. When working with 2ups intersecting frames are indicated by red circles:

The distance between green and blue frame represents the margin.



... it would betray great
I not aware of the adva
these Reports to the world
of your avowed approba
Hence, Sir, I consider th
our good opinion as an h
alf gratulation, and the testi
ent as a fute recommendati

Change the size of a frame by dragging the line when the cursor changes to arrows



or rotation angle appears when moving to mouse near an edge. Drag the frame to correct position.



3.4.3.1 How to use the Prepare cropping task (basic)

In this step you have to check the content area (red frames around the printed area) and to define the final page size (blue frame). Additionally here you deskew the page. These two steps can be made in one. Be careful: all content needs to be in the red frame (plus the margins), otherwise it will be cut and lost.

Use basic cropping to crop images, clean the borders or split 2ups.

Adjust the blue frame so that the desired information is inside.

All pages are set automatically to individual page type.

Use the 'Preview final page' button to see how the page will look like after cropping

You can sort columns by clicking on the top row to see f.e. pages with smallest or largest height or width and correct sizes in left hand image view, if required.

If you select one or more items, the selected pages will be shown in the image view

Use (Shift) and the cursor will change to a hand. Hold left mouse to move the selected page frame without changing the frame size.

Shortcuts for faster and effective cropping

- (Ctrl+E) shrink current frame by 2 mm
- (Ctrl+L) select left frame only
- (Ctrl+R) select right frame only
- (L) rotate left (0.1 degree)
- (R) rotate right (0.1 degree)
- (G) show/hide grid
- (Shift + Tab) check previous

3.4.3.2 How to use the Prepare cropping task (advanced)

Advance cropping is used to crop images, clean the borders or split 2ups with enhanced tools.

Adjust the red frames so that the desired information is inside.

Use the Preview final page button to see how the page will look like after cropping

You can sort columns by clicking on the top row to see f.e. pages with smallest or largest height or width and correct sizes in left hand image view, if required.

If you select one or more items, the selected pages will be shown in the image view.

When creating a new frame: If you replace an existing frame the angle and the frame type of the previous frame are kept, else the angle of the new frame will be 0 and frame type will be regular

Move selected frame without size changing - use (Shift) and the cursor will change to a hand -hold left mouse. This works for regular, individual and alternative page frames.

Shortcuts for faster and effective cropping

- (Ctrl+E) shrink current frame by 2 mm
- (Ctrl+L) select left frame only
- (Ctrl+R) select right frame only
- (L) rotate left (0.1 degree)
- (R) rotate right (0.1 degree)
- (G) show/hide grid
- (Shift + Tab) check previous

Set the margins (distance between the red frame and where the page will be cropped). Depending on configuration, the margins will be filled with white, black or background.

Shortcuts

- (X) remove alignment (reset to 0, final page will be centered to red frame)
- (7) align top left
- (8) align top center
- (9) align top right
- (4) previous page
- (5) align center
- (6) next page
- (2) align bottom left
- (1) align bottom center
- (3) align bottom right
- (0) shrink frame to black pixels; hide frame

Depending on configuration, the margins will be filled with white, black or background.

Use context menu (click with right mouse on a frame line) to open Page frame element and edit position, dimensions, rotation or page type.

Page frame element ✕

Position

X Y

Dimension

Width

Height

Rotation angle

Regular page size

Alternative page size

Individual page size

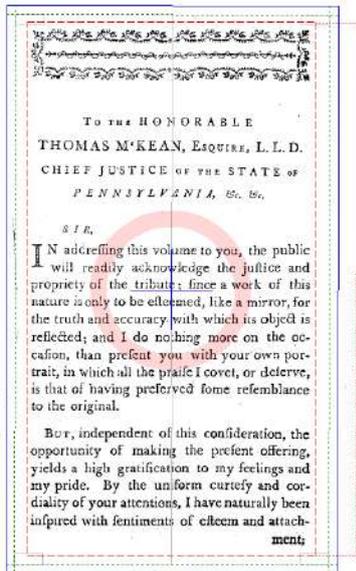
Make final adjustments (if needed).

See **Special tools for advanced cropping** chapter.

The default page width and height is set in order to define the default page size of the final (single) pages. Larger foldouts such as maps or charts will not be effected by this and remain in their original size.

Frame colors

In addition you can see the position, and how the pages will be adjusted on final pages:



 A **blue cross** shows the center of the page.

A **blue frame** shows the final page with additional margins to left, right.

 A **circle** in the center shows the currently selected frame.

Everything outside the **red frame** will be cut and filled with background color (or pattern - depending on configuration settings) during further processing steps.

 A **blue filled circle** is shown with the final page.

 The **red dotted line** is used to resize the red frame width or height by drag\$drop.

A **green frame** shows the ideal content area. Here some small rectangles are helpers for exactly positioning for example page numbers.

Ideally red and green frame lie on top of one another.

 A **red small circle** appears, if you work with 2ups and the red frames intersect.

The distance between green and blue frame represents the margin.

Move page frames

Page frames can be moved by holding down the (Shift) key before pressing the left mouse button.

A simple click moves the edge of the frame to the position where you clicked.

A function similar to manual deskew is available. Move cursor to a frame edge until it changes to a double arrow sign . Adjust frame angle, release cursor and the deskewed frame is visible.



Deskew angle attribute of page is stored as well as the angle used for applying page frames (align, 2 ups split).

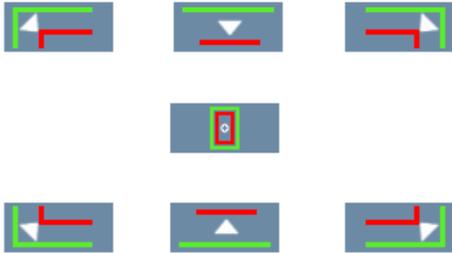
Page frames may have an angle of approx. 90, 180 or 270 degree. This will result in rotation of the page. An arrow for zones shows the orientation if not portrait. If no blue arrow is shown, it means the rotation is less than 45 degrees to the left or right.

The angle function in **Prepare cropping** only allows two digits after the decimal (45.00) when entering in context menu. When entering "45.0099" it translates to "45.00". Just two digits are allowed for precision. Internally, deskew is stored with two digits as well.

Note: There is no possibility to create polygonal zones in Prepare cropping.

Align frames automatically

First the green frame must be shown . To align frames automatically use the  automatic align tool. Use one of the alignment buttons:



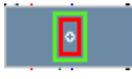
When one button is pressed, the dialog is closed and the action starts (no need to select the option and then press OK).

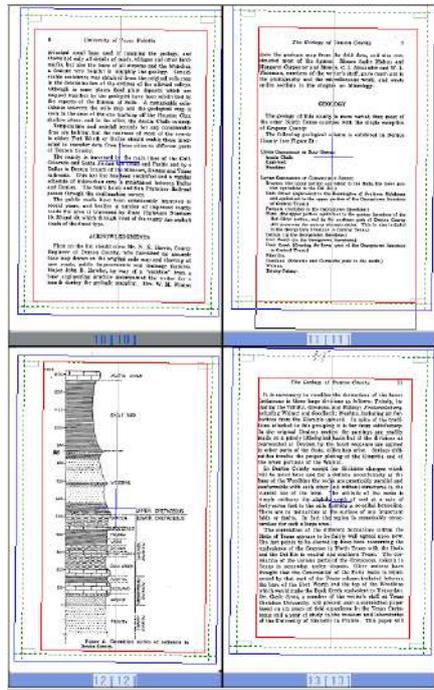
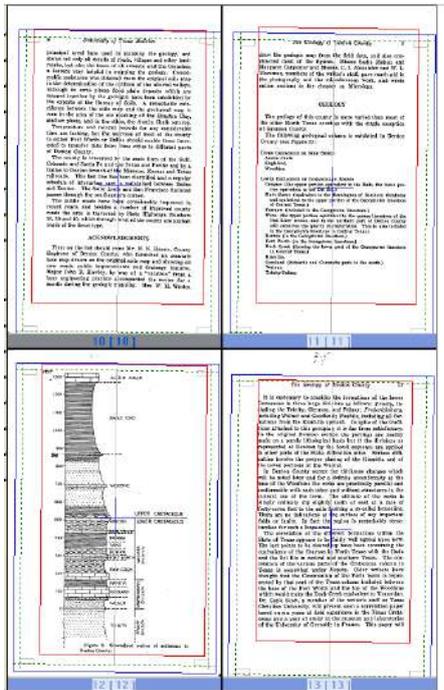
To exit automatic align tool, press (Esc) key or right click.

Alignment is done on all pages of the document and shrinks the frame rectangle using the same code like in the case of the manual alignment possibilities with keys 7,8,9,5,1,2,3.

Example:

Show content area (green frame) and final page frame (blue frame):

Click alignment tool  and the result is:



If the page frame is bigger in size than the final page (oversize) no alignment is made.

If there is a 400 px difference between the union rectangle and the shrink the frame rectangle no align is made.

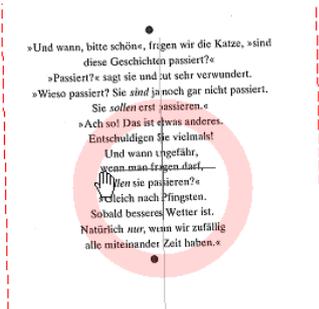
If there is a 300 px difference between the union rectangle and the shrink the frame rectangle no align is made (only if not center).

If top difference between the union rectangle and the shrink the frame rectangle is greater than 300 px and the align option is TopLeft, Top, TopRight then no align is made.

Move selected frame without size changing manually

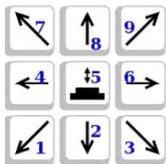
Select a frame and place mouse inside the circle. Use (Shift) and the cursor will change to a hand. Hold left mouse and you are able to move the selected page frame without changing the frame size.

This works for regular, individual and alternative page frames.



Use keyboard

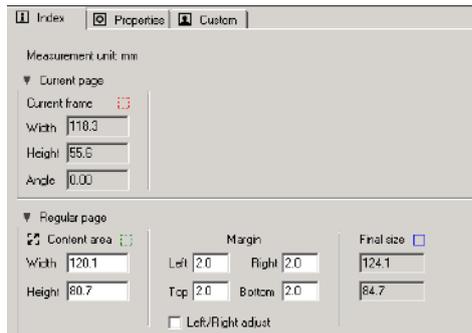
To move the selected frame and work fast, you may use the numeric keyboard (num lock shall be on):



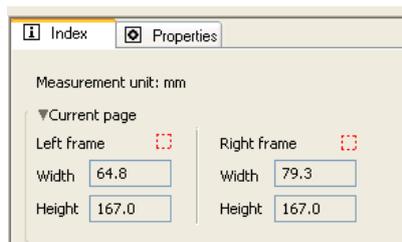
Key	Action
(x)	remove alignment (reset to 0, final page will be centered to red frame)
(7)	align top left
(8)	align top center
(9)	align top right
(4)	previous page
(5)	align center
(6)	next page
(1)	align bottom left
(2)	align bottom center
(3)	align bottom right
(0)	shrink frame to black pixels inside frame

When creating a new frame: If you replace an existing frame, the angle and the frame type of the previous frame is kept, else the angle of the new frame will be 0 and frame type will be regular.

The advanced Index view on the right hand side shows content area, final page size and other properties in collapsible groups.



Processing double pages the current page part will show left and right frames.



Regular page



Used if most of the pages have regular page sizes, and some, for example alternative pages, these can have a different page size. **Using Add Margins** adds the defined margins to individual page sizes.

The position of the frame on the page (left or right) is detected and the columns are filled accordingly.

Width and Height of the content area content area is the default printed area and default size for all pages.

Margins identify the maximum size of an additional margin in mm to the left or right, top or bottom. Those are the distances from the content area to the borders of the final pages. Value can be set in the *.rdy file.

Final size fields are computed by the other values.

Left/Right Adjust to use inner and outer margins instead of left and right. Used for asymmetric margins e.g. the left margin has more width than the right one. This setting affects the alternative pages, too.

Turn off margins

The screenshot shows four panels of settings for 'Regular page' and 'Alternative page'. The 'Regular page' panels have 'Turn off margins' unchecked, with margins set to 12.0 (Left, Right, Top, Bottom) and final sizes of 84.0 and 144.0. The 'Alternative page' panels have margins set to 8.0 (Left, Right, Top, Bottom) and final sizes of 76.0 and 136.0. In the bottom-most 'Regular page' panel, 'Turn off margins' is checked, and all margin fields are set to 0.0, with final sizes of 60.0 and 120.0. A blue arrow points from the 'Bottom' margin field in the second 'Alternative page' panel down to the 'Turn off margins' checkbox in the bottom-most 'Regular page' panel.

Turn off margins will disable the margins and set them “0.0”. This action will be applied only on the current document – other documents will have the margins set in the settings file.

Lock/unlock content area

For print-space there is a possibility to be auto-computed (unlocked ) - always take the biggest width and height for it taking in to account all the regular (red) frames. The same is true for the alternative print-space using just the biggest width alternative frame and biggest height alternative page. The auto mode is used also if you don't want to set the content area, but you want to be sure that all the cropping will fit nicely in the resulting page.

If you lock the content area  , the values (that are seen in interface) are not modified anymore - consider that the perfect content area size was chosen or will be entered by hand, later.

The lock doesn't influence the alignment (position), just the size of the content area.

Alternative page

The screenshot shows the 'Alternative page' settings panel. It includes a 'Content area' section with 'Width' set to 60.0 and 'Height' set to 120.0. The 'Margin' section has 'Left', 'Right', 'Top', and 'Bottom' all set to 0.0. The 'Final size' section has 'Final size' set to 60.0 and 'Height' set to 120.0. The 'Final size' checkbox is unchecked.

An alternative page can be a supplement or a foldout or a page in a different size than most of the other pages.

Width and Height of the content area is the default printed area and default size for all pages.

Margins are taken from the regular page size.

The **final size** is computed automatically from the width/height plus margins.

Page verification

This list shows all pages and their current status.

Page verification									
<input checked="" type="checkbox"/> Use list for pages order									
No.	Page	Type	Status	Res.	Color	Width	Height	Width2	Height2
1	right	<input type="checkbox"/>	ok	300	gray	118.3	55.6		
2	left	<input type="checkbox"/>	align error	300	gray	117.8	72.2		
3	right	<input type="checkbox"/>	align error	300	gray	117.1	72.5		
4	left	<input type="checkbox"/>	ok	300	gray	120.1	80.7		
5	right	<input type="checkbox"/>	align error	300	gray	118.0	71.9		
6	double	<input type="checkbox"/> <input type="checkbox"/>	ok	300	gray	54.9	72.0	48.5	46.4
7	right	<input type="checkbox"/>	align error	300	gray	117.5	71.5		
8	left	<input type="checkbox"/>	align error	300	gray	118.1	72.2		
9	right	<input type="checkbox"/>	align error	300	gray	117.6	71.2		
10	left	<input type="checkbox"/>	align error	300	gray	118.6	71.9		

Description of the columns:

No.	Sequential number
Page	Page number
Align	Contain as possible values: double, left, right, single, cover, spine, edge, foldout. Those are marking the page type (align) (same like on right click in the page control)
Type	Column contains the possible values: R, I, A, “ “ or any combination of these values in case of double frame page (for example R/R). The short names come from the frame types: Regular, Individual, Alternative page or empty in case of missing frames.
Status	<p>Ok: page status is correct, no problems detected.</p> <p>Empty: no frames on that page.</p> <p>Suspicious: the detected content area (green frame) exceeds the final page size. Check red frames on these pages again. (width or height plus a tolerance of 4.5 mm)</p> <p>Oversize: the frame is detected as oversized (much larger than the final page (blue))</p> <p>Align error: right and left hand pages are not in the correct order. Green frame exceeds final page size or red frame exceeds green/blue frame size. (Width plus a tolerance of 4.5 mm and height plus a tolerance of 4.5 mm)</p> <p>Also appears on double frame pages if the content areas are not vertically similar. (Difference between the vertical content area center is larger than 15 mm)</p> <p>Error: More than 2 frames on page or more than one frame on pages that are not set as a double frame page.</p> <p>Missing frames: left or right frame is not set in the case of 2 ups.</p>
Res.	Image resolution in dpi
Color	b&w, color or gray
Width / Height	Dimensions of the regular page frame size (red frame)
Width2 / Height2	<p>Dimensions of the second (red) frame in case of double pages.</p> <p>In the image above, on page 9 the right frame is missing, and on page 10, the left frame is missing.</p>

You may click on the column headers to sort the list. Check **Use list for pages order** to show the page images on left hand side image view in same order as shown here in the page verification list.

Special tools for advance cropping

Some cropping tools are only available in advanced cropping and are hidden when working with basic cropping mode.



Preview final page

Shows the background in gray so that the final page size can be viewed better and mistakes can be seen immediately.

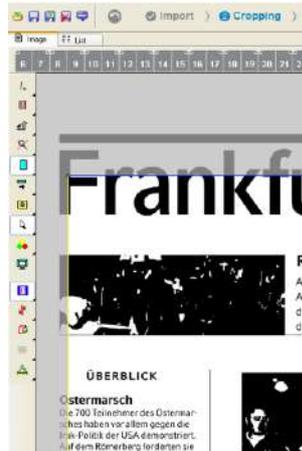


Image on the left shows the Preview final page button pressed, the image on the right is without it.

You can better see that parts of the text are not included in the final page size frame.



Show content area - green frame

This tool is only active for advanced cropping mode. Guiding frame to display content area (green) on each page for advanced alignment.



Show final page - blue frame

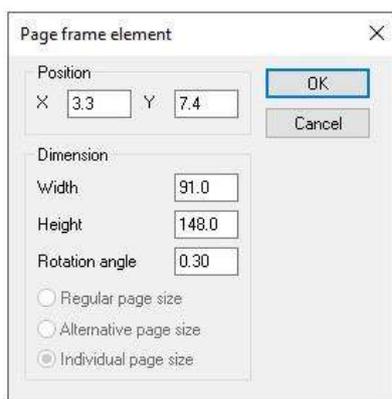
This tool is only active in advanced cropping mode.

The view of page spaces can be turned on and off by **Show content area**. When it is turned on, you may move the position of the page space using drag & drop inside the page frame.

First, raw scans include a lot of background which must be removed using crop algorithms. When applying a crop process, the content area and some margin must not be cut off. Specifically no text must be cut off. This must be verified manually as it is the most important task to provide excellent reproductions of the bindings replicating as close as possible the original page and book in digital form.

In case the automated algorithms do cut off text, an operator can manually change the crop frame which is typically set around the content area including additional margins. Also, the rotation can be changed if it has not been detected correctly, so that the final image looks perfect. You can use cursor keys to set the right position.

Use right mouse click to open the page frame element mask.



Content area auto-computes and has the possibility to lock Content area size by  tool.



Set content area sizes from current frame

Only available for advanced cropping mode.

Changes the width and height of the content area as the width and height of the current frame; if the frame selected is regular or individual, the content area width and height are changed;

If the frame selected is an alternative page, the alternative page content area width and height are changed.



Set best content area for regular page size

Selects from the regular frames sizes the biggest width and height and sets the content area at these sizes; Alternative page and individual frames are not used when selecting the biggest sizes.



Set best content area for alternative page size

Selects from all the alternative page frame sizes the biggest width and height and sets the content area at these sizes;

Regular and individual frames are not used when selecting the biggest sizes.



Set content area width from current frame

Sets the width of the current frame as the width of the content area.

If selected frame is regular, the width of the content area is changed.

If selected frame is individual, no sizes are changed.

If selected frame is alternative page, the width of the alternative page content area is changed.



Set content area height from current frame

Sets the height of the current frame as the height of the content area.

If selected frame is regular, the height of the content area is changed.

If selected frame is individual, no sizes are changed

If selected frame is alternative page, the height of the alternative page content area is changed.

Note: "Set content area width from current frame" and "Set content area height from current frame" apply only on the type of the current frame (if current frame is alternative page, the actions change the alternative page content area; if the current frame is regular, the actions change the regular content area).



Copy only frame size - all images (keeping center and angle)

This helps in cases, where on many pages the detection failed for some reason (for example isolated page number at bottom not recognized). Enlarge to current allows different sizes for left and right frame on double pages.



Copy only frame angle - all images

This action works in the same way as "Copy only frame size to all images": it sets the angle of the current frame to all frames of the same type.



Copy only frame position - from current page to end

This action will move the frames to the top-left position of the current frame. It applies on certain type of frames, depending on the type of the current frame and the type of the page (left hand side or right hand side). At the end of the action, a message will appear announcing the user whether the action was successful or not and which frames the action was applied on.

If current frame is regular frame on left hand side, the action will apply on all regular frames from left hand side pages, starting with the next left hand side page that has a regular frame (the same for regular frame on right hand side page)

If current frame is individual frame on left hand side, the action will apply on all individual frames from left hand side pages, starting with the next left hand side page that has an individual frame (the same for individual frame on right hand side page)

If current frame is alternative page frame on left hand side, the action will apply on all alternative page frames from left hand side pages, starting with the next left hand side page that has a alternative page frame (the same for alternative page frame on right hand side page)

Action is not available for 2 up pages.



Copy frame size, position and angle - all images

Copies size, position and angle of the current frame to all similar page frames.



Open Shrink / Enlarge dialog

Opens a dialog in the center of the Image View, in which all actions needed to change the frame sizes are available. This tool is only available for the advanced cropping mode.



Automatic shrink all frames

Shrinks frames as much as possible. This tool is only available for the advanced cropping mode.



Regular page size

Only available in advanced cropping mode.

Press "regular" to mark pages with normal dimensions. The majority of the document should have pages with regular, consistent page sizes. In exceptional cases (foldout, manuscripts, others) use Alternative page or Individual page types. By default, all pages shall get the same size after processing.



Alternative page size

Press "Alternative page size" to mark a page as an alternative page. Alternative pages can have a different page size from normal pages.

Example: A daily newspaper contains each Friday the TV program as an alternative page with a different size. Using this tool - the color of the frame will change to green. Only available in advanced cropping mode.



Individual page size

This is the default value. Use this to mark an individual page type. Used for maps, foldouts or particular pages.

With a right click the toolbar will expand with more options. Example: a map in a book which is folded to half the size of the book. The fold out will have a larger width but the same height. Using this option the color of the frame will change to blue. Only available in advanced cropping mode.



Set to final size

Increases the frame to the size defined in the "Final Page Size" step from the Index view and changes the frame automatically (no matter the frame type).

This is useful on colored cover pages as these are usually the largest pages of the document. After usage, the red and blue frame will have the same size. Only available in advanced cropping mode.



Change frame type for all pages

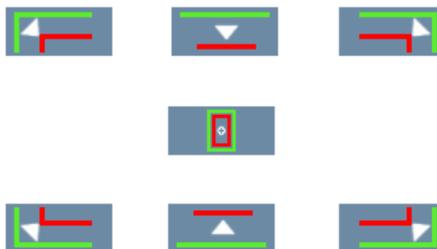
Only available for advanced cropping mode.

Chauken, die an der Nordseeküste auf künstlichen Hügeln sich so gegen die Sturmfluten zu schützen. Solche Hügel, die der Warfen, ziehen sich an der ganzen Küste vom Rhein bis zur renze entlang. Ihre Anfänge stammen meist aus den letzten vor-Jahrhunderten, als das Küstenland zu sinken begann. Im Küsten-Nordsee sind im Lauf der Jahrtausende mehrere Senkungen und ad worden. Als in en Periode der Erde nicht me n das Wohnen auf der Erde nicht me die Anwöner, vor allem die Chauken und die Friesen gezwungen, Irige, dann aber immer höhere Hügel zu erbauen, bis schließlich d Christi Geburt mit dem Bau des goldenen Riffes



Automatic Align

Only available for advanced cropping mode and in the Review cropping task. See **Align active frames manually** and **Align frames automatically**.



When one button is pressed, the dialog is closed and the action started (no need to select the option and then press OK).

To exit Automatic Align tool, press (Esc) key or right click.



Center by page numbers

Tells the system to try to find the page numbers and then choose them for centering instead of using the printed text. Only available for advanced cropping mode and in the Review cropping task.



Show grid

Switches grid on and off. Click again to get another view/size of the grid.
Shortcut: (G)

The grid can be configured for any view in any step. Outline image view (not in full screen, or detail image view).

The grid button by default shows an outline image view with a blue 10X10mm grid. The color and size of the grid can be changed.

The Grid button has 3 modes:



No grid. Grid view disabled.



Shows the grid as helper only when the frame rotation angle is 0°.



Shows the grid as helper, independent of rotation.



Show measurement

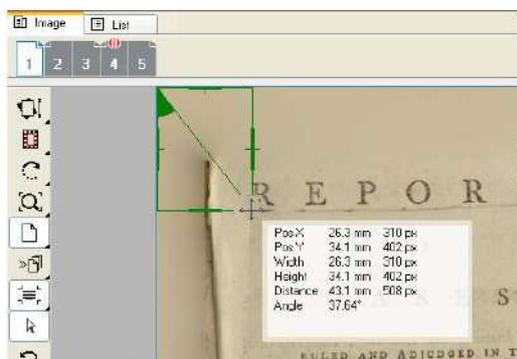
In **Prepare** and **Review Cropping** you can toggle measurement units used in all values from dimensions controls. If pressed, all values are expressed in inch, if not pressed, all values are expressed in tenth of millimeters. The default is depending on the current local configuration settings. Changing measurement type here will only change the view inside dW and will not change the measurement unit of the document itself.



Measurement tool

When opened, the first corner is always set to 0, 0 (top left corner of the page). When moving the mouse, the Width, Height, Distance and Angle measured are displayed as a green rectangle.

Move the cursor over the image on the left hand side. Hold left mouse button and move cursor to another place. Useful to measure content area sizes.



Shows position on the image in 1/10th of mm.

Also, a combobox is moved with the mouse cursor, where the values are displayed.

- Pos X and Pos Y** the current mouse position (is not changed to 0, 0 if mouse click)
- Width** distance between last point (or 0, 0 if no clicks were made) and current mouse position, on X axis
- Height** distance between last point (or 0, 0 if no clicks were made) and current mouse position, on Y axis
- Distance** direct distance between last point (or 0, 0 if no clicks were made) and current mouse position (the rectangle diagonal)
- Angle** the angle made by the rectangle diagonal and the Y axis

When mouse is clicked, the new point becomes the new 0 point and the new distances will be measured from that point (except Pos X and Pos Y). To exit the measure tool, press (Esc) key or right click.



Color picker

Shows RGB values besides the position. The color picker can be used to check if a target page is scanned properly. Leave the color picker by (Esc).





Copy page

Creates a copy of the current page. Needed for more than two frames. In case you have a 2up page (double page) in a single page document and you want to split it in single pages.

Select the page in outline view (right) in dW, in Review Import/Prepare Cropping step.

Duplicate the page using the  button.

-> A new copy of selected page will be created and will be inserted directly after the selected page. If you have selected more than one page, the first selected page will be copied. This action does not have any effect to the original files.

The copy will be created in internal document pool.

Draw a frame for first page in the current image and the second page in the next image.

Use (Delete) key to delete a frame in case there are two many detected.

Same procedure for 4 pages in a 2up scan.

In case of an upside down image use the rotation functions in the left toolbar (extended button list) to rotate the page image

3.4.3.3 How to use the Review cropping task

This task is especially designed for QA after image pre-processing. It can be skipped if analysis tasks like **Review page sequence** will follow.

In detail it works similarly to: [How to use the Prepare cropping task](#) chapter.

Note: If you work with 2-ups, click once in one of the pages, otherwise changes will be done on both pages.

The index view in **Review cropping** differs from the one in **Prepare Cropping**. This task is optimized for verification of pages.

3.4.4 Zoning

docWizz has analyzed the scanned pages and used all its tools and skills to determine not only the geometric size and dimensions of all document elements but to define, which category the zones should be sorted to

3.4.4.1 How to use the Review zoning task

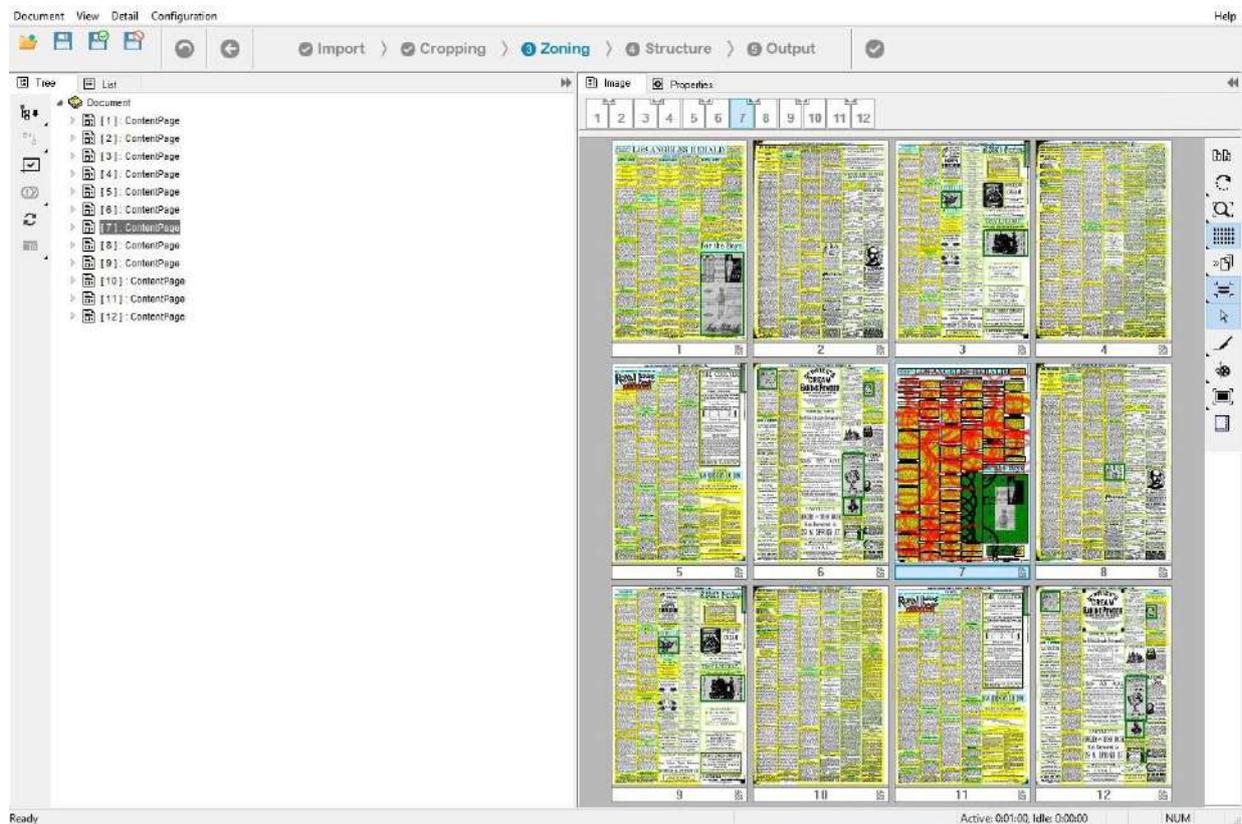
The Review zoning task is used to check if the zones and the classifications have been recognized correctly.

When accessing the task Review zoning the most time consuming process, the OCR has been done. However, OCR correction should take place later in the task Review structure and text. Also, the recognition of all parts of the document has been done. docWizz separates the document and even each page into zones. Zones can contain text, illustrations, tables etc. In a later stage, the text zones will be classified more in detail (headlines, footnotes, etc.).

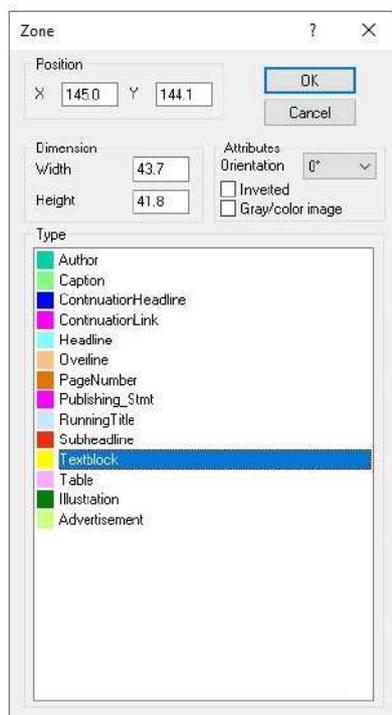
In Review zoning you check if all parts of the document have been recognized correctly so far. It is recommended to check this in a multiple page view in order to accelerate the speed of checking. Focus

on the recognition of illustrations and tables, too. docWizz has very good tools for checking and correcting the features selection of multiple zones, merging, deleting, cutting and analysis of TOC (table of contents).

The image view of the pages shows the result of the analysis. All elements are marked with different colored zones.



The color refers to the type and the element that has been automatically sorted to e.g. Advertisement, Formula, Illustration, Table or Vertical textblock.



Putting the cursor on a zone and pressing the right mouse button opens the **Zone** dialog, which shows the already available information about the current zone. You can assign a zone type to another zone by selecting it in the **Type** window and pressing .

In this step, each layout element must be classified correctly by its type. This is especially important for headlines as the logical headline hierarchy is based on the results.

Change to **Full Zoom View** by the  tool in right bar.

Check relevant zone(s). Select multiple zones by drawing a frame around these (hold left mouse button and draw frame).

Some helpful commands:

Merge: [F8] (also makes rectangular zone from polygon).

Activate Knife tool to cut: [F9] (right hand click or [Esc] to release).

Change zone type: Hit the appropriate short key (first letter of zone type: [P] for Page Number, [T] for Textblock, [TT] for Table, [A] for Author, [AA] for Advertisement, [H] for Headline, etc.).

Check each page and correct zones where necessary. Use [Page Down] to get to next page. When finished hit the green process button (top right). To flip through the batch of pages you may also use the tree view in the left window. You can call up pages as well as single zones.

Cut/Paste page(s): These actions are available in Review zoning and Review page sequence and can be used to move pages from one document to another, keeping the layout analysis and the OCR detection. The actions work only on documents from the same project configuration and with the same task. "Cut page(s)" action only marks the page(s) for moving. Closing docWizz or performing other actions will not make the cut pages disappear from the document.

3.4.4.1.1 Correct zones in Tree view

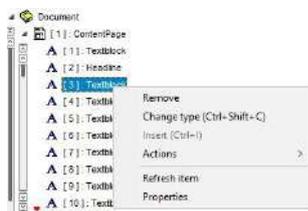
If you notice zones which have not correctly been traced, you can correct them by executing the following description.

Assume docWizz has missed marking the headline as a separate zone.



Put the cursor on the icon of a current page item such as a text block in the tree view in the left window and press the right mouse button.

Pressing the right mouse button, with the cursor on top of a text block entry, opens a context menu.



In this chapter we describe entries of the context menu. Which entries are available in the context menu may change from task to task and depend on whether you are working on monographs or newspapers or other document types.

Remove

Removes selected zone in list view and in image view.

Change Type

Selecting **Change Type** opens a small list window attached to the textblock icon. Select the desired type name (Advertisement, Formula, Illustration, Table, Headline or others) in the list and assign it with a double click of the left mouse button to the zone.

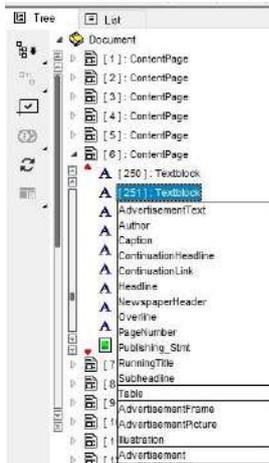
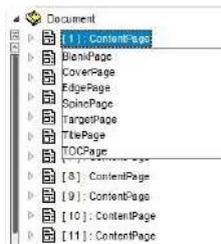


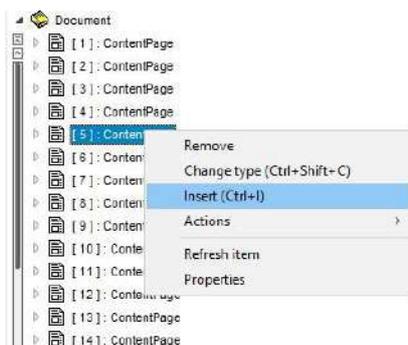
Figure 2: Change Type - List

You can also change the type of content page or other page types the same way.



Insert

Select **Insert** in the context dialog. Choose between **Missing Page** and **Page from File**.



After the system has calculated the new entry and has integrated it into the context of the document, the new zone is highlighted in yellow and is defined as a text zone. You may assign the zone to another zone type as described above.

Move to

Front-Main-Back: If a monograph is processed, you just deal with the movement of pages to Front, Main and Back.

docWizz separates documents into three major parts: Front, Main and Back. Front consists of elements (pages) such as title page, preface, table of contents, etc. All the content (chapters, or if a serial is processed, issues). Back may contain for example Appendixes. At this stage, only a determination between Front, Main and Back is done. The more detailed distinction takes place in the next task within docWizz.

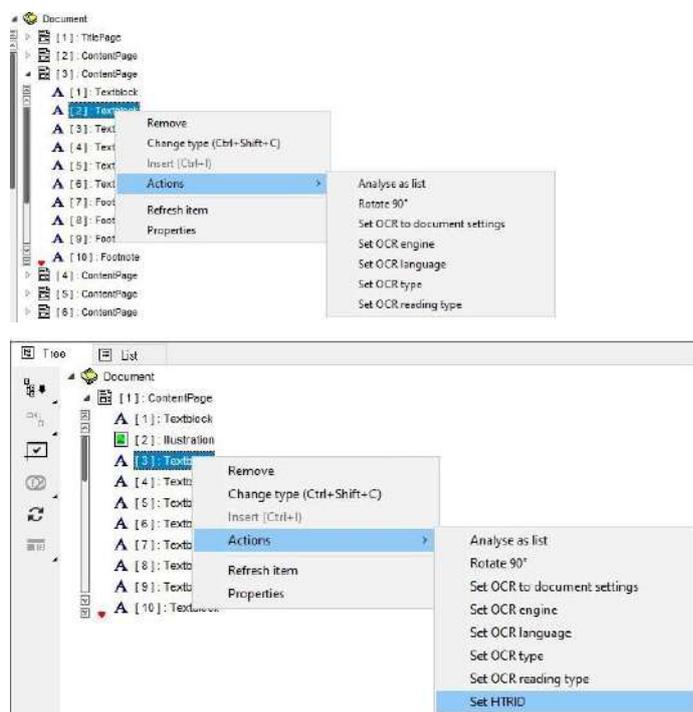
In order to easily change the results of automatic recognition, you can move a sequence of pages to a different part.

Front-Issues-Back: If the current document has been processed as a **Serial** you must check if all Issues have been detected and separated correctly. You may also verify the front and back part of the entire book. Pressing the right mouse button on the issue icon in the tree view may insert or remove issues. If you delete an issue, its content will be moved to the previous issue.

If a serial is processed, the user allocates the pages to Front, Issues and Back. Issues contain Front, Main and Back levels, too. So the determination is more complicated when working with serials.

Note: It is not possible to move multiple pages at once. The sequence shall remain unchanged. So you just select the last/first page to be moved. All pages before/after are automatically moved to the different section..

Actions for textblock



Analyse as list: If the recognition of a part in a document that looks similar to list lead to a bad result (for example too many zones), the special feature Analyze as List might provide a much better result. To use this feature, select the zone or zones belonging to the list. Then, move the cursor to the tree view (usually on the left side of the user interface) and do a right hand click on one of the highlighted entries. A context menu appears. Choose 'Actions' and here 'Analyze as TOC'. docWizz now reanalyzes the zone or zones and treats it as an object as a list. The result should be much better than the one before.

Rotate 90: Adds small blue arrows to show the rotation.

Set OCR to document settings: Set the OCR engine and version, the language and the font type of the selected zone(s) the same as the document's. The action can be used to revert all the OCR changes made on a zone.

Set OCR engine: Open a list with the supported OCR engines, from which the user can select on OCR engine to perform the OCR on selected zone(s).

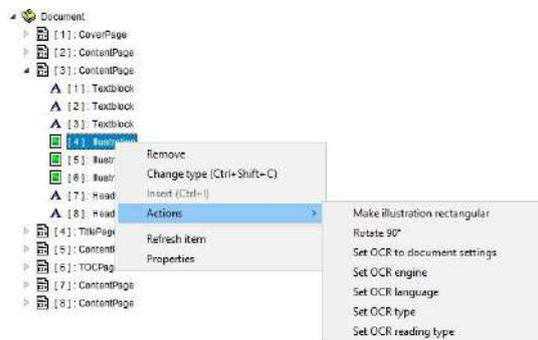
Set OCR Language: Select language from the list appearing.

Set OCR Type: Select from list Antiqua, Fraktur, Typewriter, Fraktur+Antiqua or Typewriter+Antiqua.

Set OCR reading type: Select from list Auto, Horizontal stripes (by line), Vertical stripes (by column).

Set HTRID: Select from list the available models and HTRID values. The menu entry "Set HTRID" is displayed if the selected zone has OCR Engine Transkribus. For Transkribus, since it deals with a lot of different writing styles, both handwritten or printed, language is less important, more important is the writing style, or maybe custom model for a particular collection. This is why you need to set also the ID of the model used.

Actions for illustration



Make illustration rectangular: Changes the shape of an polygonal into a rectangular shape.

Rotate 90: Adds small blue arrows to show the rotation.

Set OCR to document settings: Set the OCR engine and version, the language and the font type of the selected zone(s) the same as the document's. The action can be used to revert all the OCR changes made on a zone.

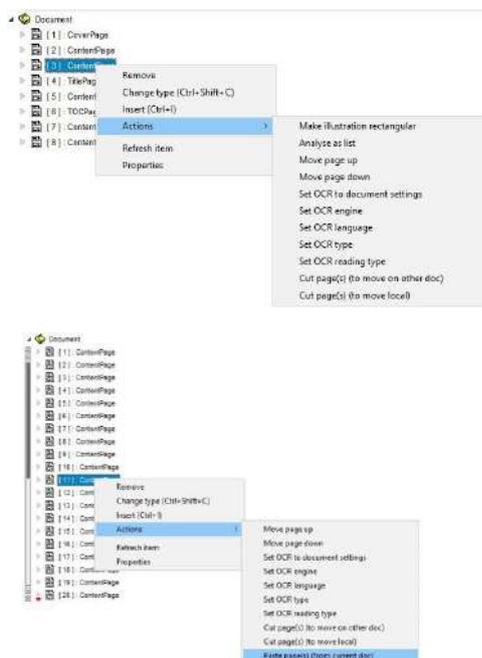
Set OCR engine: Open a list with the supported OCR engines, from which the user can select on OCR engine to perform the OCR on selected zone(s).

Set OCR Language: Select language from the list appearing.

Set OCR Type: Select from list Antiqua, Fraktur, Typewriter, Fraktur+Antiqua or Typewriter+Antiqua.

Set OCR reading type: Select from list Auto, Horizontal stripes (by line), Vertical stripes (by column).

Actions for content page



Make illustration rectangular: Changes the shape of an polygonal into a rectangular shape.

Analyse as List: If the recognition of a part in a document that looks similar to list lead to a bad result (for example too many zones), the special feature Analyze as List might provide a much better result. To use this feature, select the zone or zones belonging to the list. Then, move the cursor to the tree view (usually on the left side of the user interface) and do a right hand click on one of the highlighted entries. A context menu appears. Choose 'Actions' and here 'Analyze as TOC'. docWizz now reanalyzes the zone or zones and treats it as an object as a list. The result should be much better than the one before.

Move page up: Page is moved one page up.

Move page down: Page is moved one page down.

Set OCR to document settings: Set the OCR engine and version, the language and the font type of the selected zone(s) the same as the document's. The action can be used to revert all the OCR changes made on a zone.

Set OCR engine: Open a list with the supported OCR engines, from which the user can select on OCR engine to perform the OCR on selected zone(s).

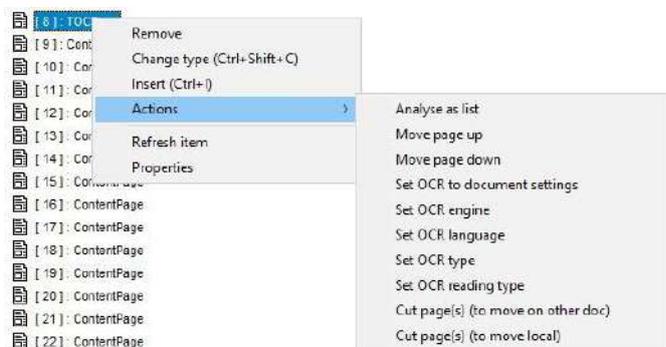
Set OCR language: Select language from the list appearing.

Set OCR type: Select from list Antiqua, Fraktur, Typewriter, Fraktur+Antiqua or Typewriter+Antiqua.

Set OCR reading type: Select from list Auto, Horizontal stripes (by line), Vertical stripes (by column).

Cut / Paste page(s): These actions can be used to move pages inside the document easily. Right click on the pages you want to move, select "Cut page(s) (to move local)". Then navigate where you want to move the pages and select "Paste page(s) (from current doc)". The pages will be inserted before the selected page.

Actions for TOC



Analyse as List: If the recognition of a part in a document that looks similar to list lead to a bad result (for example too many zones), the special feature Analyze as List might provide a much better result. To use this feature, select the zone or zones belonging to the list. Then, move the cursor to the tree view (usually on the left side of the user interface) and do a right hand click on one of the highlighted entries. A context menu appears. Choose 'Actions' and here 'Analyze as TOC'. docWizz now reanalyzes the zone or zones and treats it as an object as a list. The result should be much better than the one before

Move Page Up: Page is moved one page up.

Move Page Down: Page is moved one page down.

Set OCR to document settings: Set the OCR engine and version, the language and the font type of the selected zone(s) the same as the document's. The action can be used to revert all the OCR changes made on a zone.

Set OCR engine: Open a list with the supported OCR engines, from which the user can select on OCR engine to perform the OCR on selected zone(s).

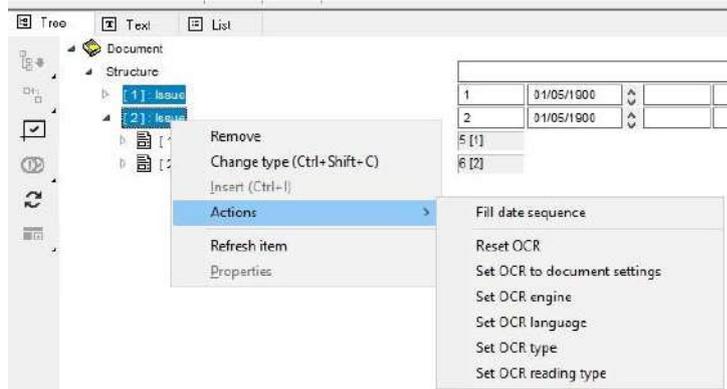
Set OCR Language: Select language from the list appearing.

Set OCR Type: Select from list Antiqua, Fraktur, Typewriter, Fraktur+Antiqua or Typewriter+Antiqua.

Cut / Paste page(s): These actions can be used to move pages inside the document easily. Right click on the pages you want to move, select "Cut page(s) (to move local)". Then navigate where you want to move the pages and select "Paste page(s) (from current doc)". The pages will be inserted before the selected page.

Fill date sequence

It is available if two or more issues are selected using (Shift) key.



You can select then the frequency of issues. Filling starts from the date of the first selected issue.

Refresh Item

If changes are not shown immediately, press **Refresh** to update the item.

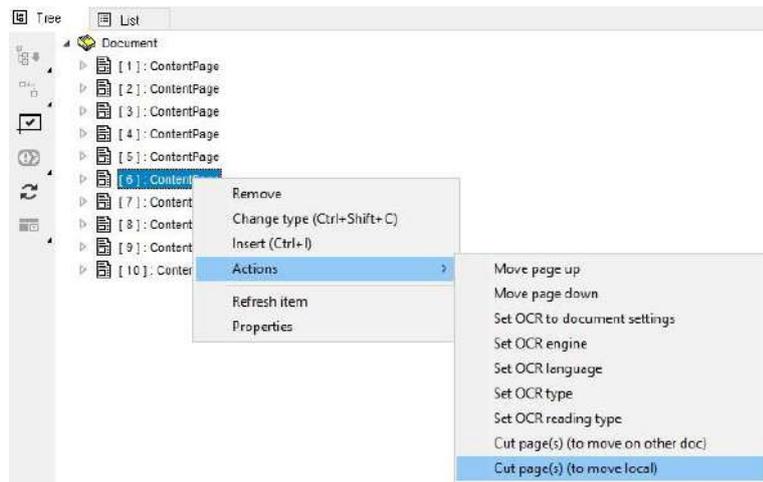
Properties

Selecting **Properties** would open the **Properties** dialog that shows information about the current text block.

Cut / paste page(s)

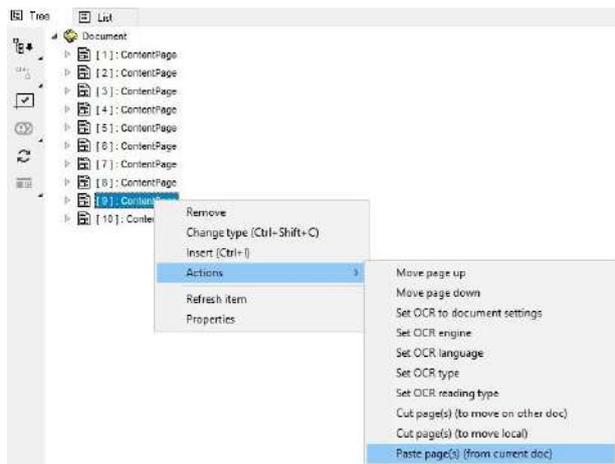
These actions can be used to move pages inside the document easily.

Right click on the pages you want to move, select "Cut page(s) (to move local)".



Then navigate where you want to move the pages and select "Paste page(s) (from current doc)".

The pages will be inserted before the selected page.

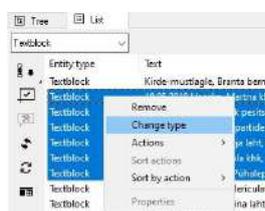


3.4.4.1.2 Correct zones in List view

Assumed docWizz has missed to mark headlines as a separate zone. First reduce the zone to the textblock.



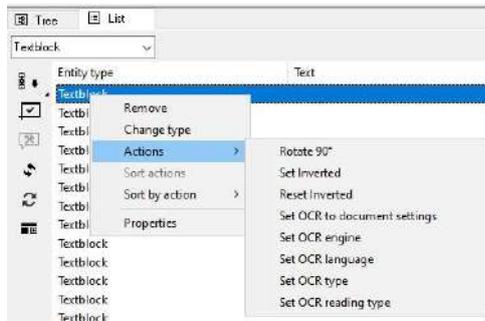
Then select the desired new type from the list (Advertisement, Formula, Illustration, Table, Headline or others).



It is also possible to change type of multiple selected types at once.

Like known from other Windows programs use (Ctrl) and click or (Shift) and click to select multiple textblocks.

Actions for Textblocks



Rotate 90: Adds small blue arrows to show the rotation.

Set Inverted / Reset inverted: Set (or reset) the flag for inverted on the selected zone(s).

Set OCR to document settings: Set the OCR engine and version, the language and the font type of the selected zone(s) the same as the document's. The action can be used to revert all the OCR changes made on a zone.

Set OCR engine: Open a list with the supported OCR engines, from which the user can select on OCR engine to perform the OCR on selected zone(s).

Set OCR Language: Select language from the list appearing.

Set OCR Type: Select from list Antiqua, Fraktur, Typewriter, Fraktur+Antiqua or Typewriter+Antiqua.

Set OCR reading type: Select from list Auto, Horizontal stripes (by line), Vertical stripes (by column).

Set HTRID: Select from list the available models and HTRID values. The menu entry "Set HTRID" is displayed if the selected zone has OCR Engine Transkribus. For Transkribus, since it deals with a lot of different writing styles, both handwritten or printed, language is less important, more important is the writing style, or maybe custom model for a particular collection. This is why you need to set also the ID of the model used.

3.4.4.1.3 Correct zones in Image view

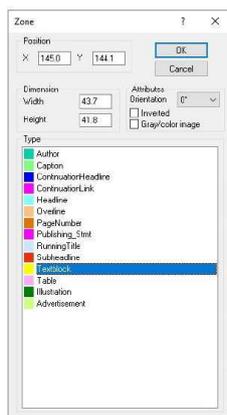
In docWizz, zone is a common expression. Elements on a page such as a headline, a text block, an illustration or a caption are recognized automatically and highlighted with a frame, what we call zone.

Select

To select a zone, just click at the zone with a left hand click.

Change

To change zones in image view elect a zone and click with right mouse button.



Using the tools in the right toolbar you can merge or cut zones or adapt them to the printed text.

When moving the mouse over a zone, its class name is shown in the page windows status bar:



Select all zones by space bar. If you are in full screen mode and move then the mouse over a zone the margins of a zone are shown, too. Zone change changes the whole page. Mouse over feature does only work in full screen mode, not in normal view.

Delete

To delete a zone, select it and press the (Del) key.

New zone

Draw a frame in the image view and select a zone type in the zone window. An entry in the tree is built automatically.

Merge

To merge zones, click at the related icon or press the (F8) key on the keyboard.

Multiple Selection

To select multiple zones, hold the (Shift) key and draw a zone. To draw a zone, press the left mouse button and hold it. Then move the mouse and draw the zone as far as you want. Lift the mouse button and every zone that has been touched by the just drawn one is highlighted

Merge: Once several zones are selected, the user can merge them with the (F8) key or the Merge zones tool.

Delete: Once several zones are selected, the user can delete them by pressing the (Del) key.

Resize

To enlarge or downsize a zone, move the cursor to a border of a zone. The cursor will change into a line with arrows at both ends. Now press the left mouse button and drag the border as far as you want.

Cut

To cut the zone afterwards, click on the knife icon in the tool bar. Then, move the cursor to the horizontal or vertical (press shift to do a vertical cut) line where you want to cut the zone. Press the left mouse button and a line appears. You can move that line. If the line is in the correct position, lift the mouse button and docWizz cuts the zone at that line.

Horizontal: Usually docWizz provides a tool to cut horizontal. To use it, click at the corresponding icon.

Vertical: Holding the [Shift] key enables the user to cut vertically.

Polygonal

To change a zone from rectangle to a polygon, move the cursor on a border line of the selected zone and right click. A new point is defined and the zone can be changed into a polygon. To erase such a point, right click on it again.

Make Polygon rectangular: If an illustration has been detected incorrect or incomplete and some noise areas in the margin were not included the operator can make the polygon rectangular. To do so, select an illustration in the tree view and right click. Choose 'Actions', 'Make Illustration Rectangular'.

Change type but do not remove block from logical structure

On changing types of blocks by pressing the first character of the word ("T" for textblock) you can switch through any types.

In case you use "R" or "P" you also change the zone type to "Running title" or "Publishing statement". When this happens the current selected elements are removed from the logical structure, as these zone types do not require logical structure. All the removed elements have to be re-assigned by drag&drop to the logical structure again. Position for the elements will be kept and just be disabled. On changing back to a type which is part of logical structure, the same position can be re-used.

3.4.4.1.4 Fast correction of zones



Fast correction mode works in full screen mode only and on steps like Review zoning or Review page sequence (steps higher than Review cropping).

For monographs and serials this feature is used in **Structure:Review issues** task.

For newspapers this feature is used in **Zoning:Review zoning**.

If enabled, this is how docWizz will behave:

- [F4]: switch to/from Fast correction
- Simple left button selection (draw rectangle over the desired zones) = merge zones, click [Esc] to cancel merge
- Simple left button click = cut zone horizontally (if the mouse is over a zone the zone type can be changed with first letter shortcuts)
- [Ctrl+R] rotates zone right (works in both Fast Correction mode and outside)
- [Ctrl+I] invert (works only in Fast Correction mode)
- When no zone is selected, only border color of zones are shown. Now when clicking into any zone, ALL zones will be shown with full color background (as per (Space) selection).
- Single right mouse click on zone deletes zone.
- Polygonal/rectangular zone: Instead of pressing (F8) twice, you can just click#drag once inside the polygonal zone and it will be turned into a rectangular zone.

Fast correction mode works in full screen mode only and on steps like Review Zoning or Page Sequence (steps higher than review Cropping).

Switch to full screen mode by the  **Full screen** tool (toolbar on the right hand side). Click the **Fast correct zones** tool  or (F4) key to switch on/off the fast correction mode. The mouse cursor changes from a cross  to an arrow .

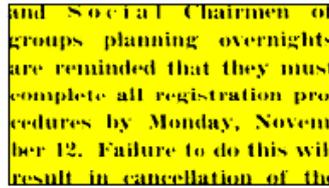
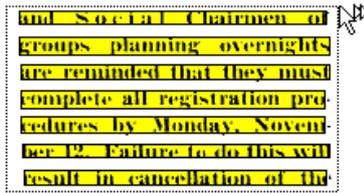
Merge zones

Separate zones can be merged using the fast correction tool and a simple left mouse button selection.

Click (Esc) to cancel merge mode.

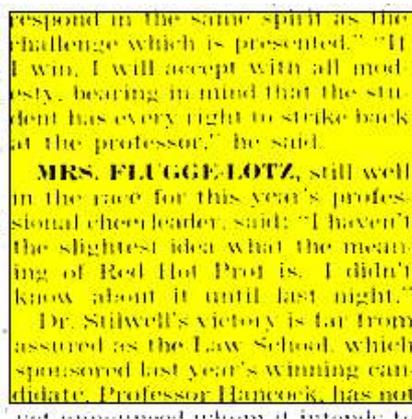
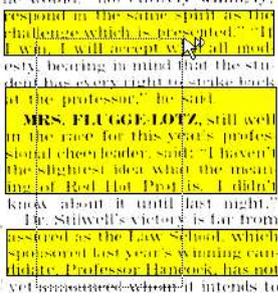
Drag rectangle over the desired zones:

The zones will be immediately merged to one zone:



Drag rectangle over the desired zones:

The zones will be immediately merged to one zone:

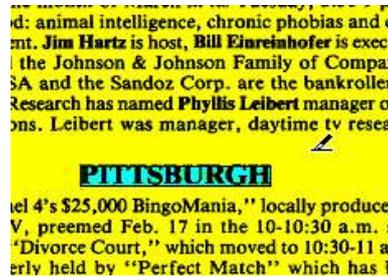
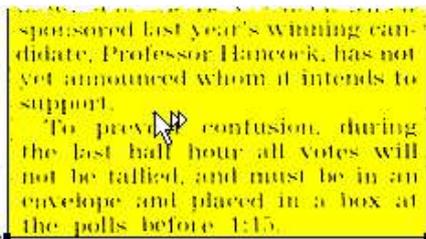


Merging zones of different zone types will merge to the "bigger" zone type.

If fast correction mode is off, zones can be merged by [F8] key.

Cut zones horizontally

Simple left button click cuts a zone horizontally if the mouse is over a zone.

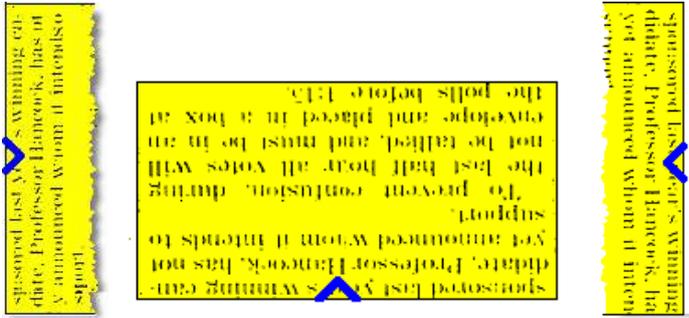


If fast correction mode is off, zones can be cut using [F9] key.

Rotate zone right

(Ctrl+R) rotates the zone 90 degrees to the right. The blue sign switches from left to bottom and right hand side if you click several times.

Works in both Fast Correction mode and outside.

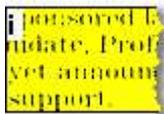


You will see a blue triangle on the side of the image which is marked as top. In export the image will either be rotated or the orientation tag is set.

With orientation the OCR recognition will be better executed.

Invert zone

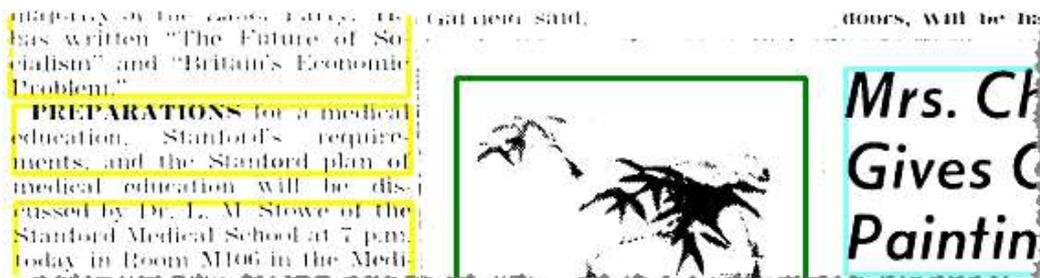
[Ctrl+I] inverts a zone. The zone gets an "i" on the top left edge. This works only in Fast Correction mode.



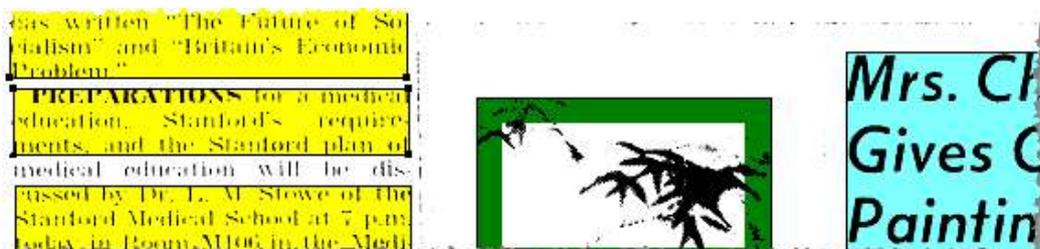
You can use the Invert button to invert an active zone and reverse the tone values. This means that black is converted to white and vice-versa. You might want to use this option when you have a source document with areas with white text on a black background, used mainly to emphasize a passage. These zones cannot be processed for automatic text recognition unless they are inverted first.

Full color background to all zones

When no zone is selected, only border color of the zones are shown.



When clicking into any zone or by (SPACE) key, all zones will be shown with full color background.

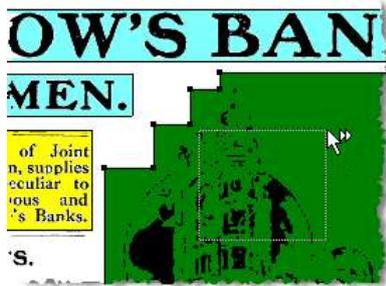


Delete zone

Single right mouse click on a zone deletes the zone.

Polygonal/rectangular zone

Just drag twice on the polygonal zone only as you would do merge for two zones:



The polygonal zone is turned into a rectangular zone:



Alternatively click twice [F8] key to get polygonal zone shape:



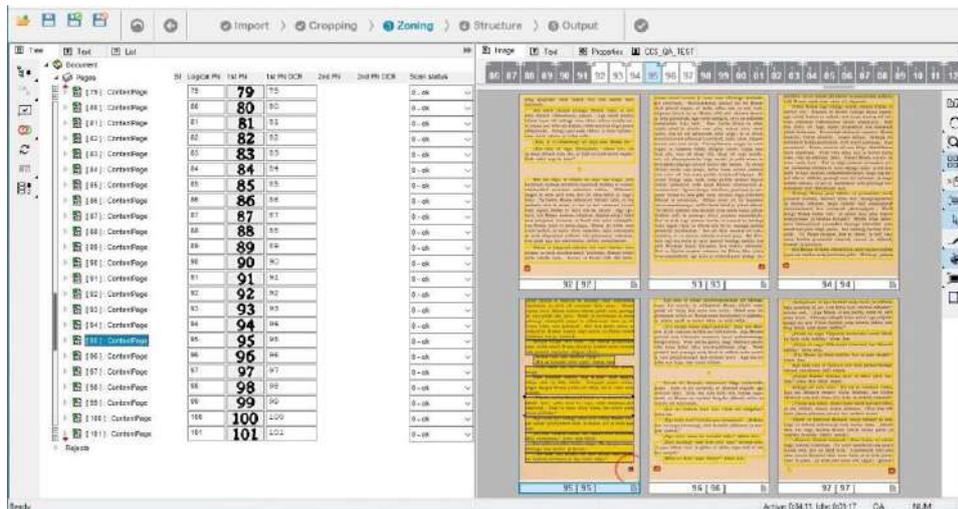
3.4.4.2 How to use the Review page sequence task

In this task you:

- check page sequence
- look for missing / duplicated pages (rescan)
- indicate issues/ sub documents
- check / Set feature codes
- check zones

Note: Classification of zones is performed in Review Issues task.

Review Page Sequence is a very simple task. Here the image page linking is performed. Basically docWizz performs the image page linking automatically and you can concentrate on the 'unsure' (unsure in terms of possible errors) elements. Therefore, docWizz identifies the zones containing the page numbers and reads the result by using OCR. In addition to this, it creates a logical page sequence and is able to fill missing page numbers automatically.



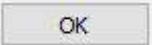
The working windows have been arranged, as shown above, to display the document structure in the tree view in the left window and miniature images of the source pages – organized by use of the  **Display Multiple Pages** button – in the right window.

In the current task you can check the book for the correct order of pages, missing pages or pages that might have been scanned twice. Use  **Previous Error** and  **Next Error** to jump directly to suspicious or flawed pages to verify and correct the systems decisions.

A detailed description of page numbers correction see **Page numbers list**.

Change zone into page number in image view

If docWizz indicated the page number as a zone already, but highlighted it as a 'text block', you can change the zone type very easy.

Simply move the cursor to the zone that contains the page numbers and do a right hand click. The Zone dialog opens. Choose **Page Number** as zone type and confirm the changes by clicking the  button. The entries for column one and three in the list on the left side should be updated manually.

Add page numbers in image view

To add page numbers, simply draw a zone around the identified page number on the original image. Change the zone type into **PageNumber**. The entries for column one and three in the list on the left side should be updated manually.

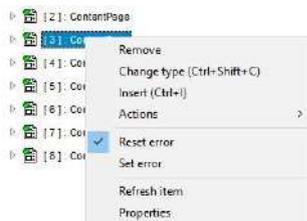
Work with the page numbers list

To verify the page numbers, docWizz performs a list with several columns.



- (1) Shows the logical page number as counted by the system and might show entries where no page number has been recognized on the page.
- (2) Shows the part of the original page image containing the page number. This part may contain text (like "page" or others). That is not used for page counting.
- (3) Shows OCR result from the original page. It may be used to correct the result of the automatic recognition.
- (4) Shows the original page image containing the second page number if working with double pages or special projects. This part may contain text (like "page" or others). That is not used for page counting.
- (5) Shows OCR result from the original page for the second page number. It may be used to correct the result of the automatic recognition.
- (6) This column shows page errors.

Click with right mouse button on **ContentPage** to open a submenu.



Remove

Removes selected item from the list view.

Change Type

To change a type of an item, just right click on the related item in the List View and double click at the desired zone type. Alternatively, you can use shortcuts

('h' for headline, 'a' for author, 'f' for footnote, 't' for table or text block (press 't' twice)). You're always able to merge two or more zones in the right part of the screen, for example if one headline has been split into two or more zones.

Insert

To build a further issue from an existing one, do a right click on the related page in the tree view and select 'Insert'. The current issue will be split beginning with the selected page. The new issue that is built contains the content from that page on up to the end of the current issue.

Move to

docWizz separates documents into three major parts: Front, Main and Back. Front consists of elements (pages) such as title page, preface, table of content, etc. All the content (chapters or, if a serial is processed, issues). Back may contain for example appendixes. At this stage, only a determination into Front, Main and Back is done. The more detailed distinction takes place in the next step within docWizz.

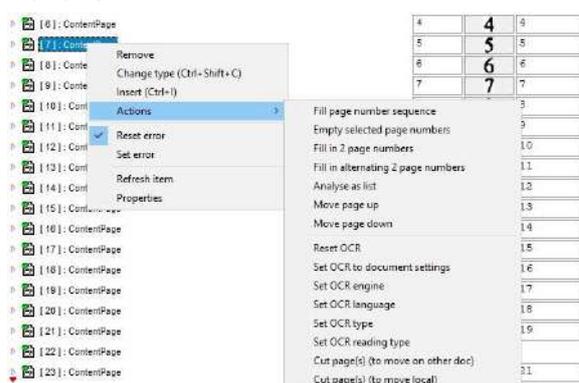
In order to easily change the results of automatic recognition, the user is able to move a sequence of pages to a different part.

Front-Main-Back: If a monograph is processed, the user just deals with the correlation of pages to Front, Main and Back.

Front-Issues-Back: If a serial is processed, the user allocates the pages to Front, Issues and Back. Issues contain Front, Main and Back levels, too. So the determination is more complicated when working with serials.

It is not possible to move multiple pages at once. The sequence shall remain unchanged. So you just select the last/first page to be moved. All pages before/after are automatically moved to the different section.

Actions



Fill page number sequence: If you have to fill some page numbers in a row, you can use the feature **Fill page number sequence**. Therefore, select the last page that contains a page number in the left column and select all following pages, which do not contain page numbers too. To do the multiple selection, hold the (Ctrl) key while clicking on each page. Then, do a right click on any highlighted page item in the tree view and choose **Actions-Fill page number sequence** in the context menu. It is possible to avoid the renumbering of page numbers even if a new issue was tagged in the serials configuration. This is due to that serials often times have continuous page numbers over the whole year and the operator has many double work to do to change back the page number series. Therefore auto-renumbering for Serial while setting issue start is disabled in the default configuration. It was before in `..\config\PVSCFG\docwizz-VSCfg.xml`
`<Propdefault name="startIssue">`

Note: Hebrew numbers are supported until 400. Page numbers starting with 400 will not be filled and no action is taken for bigger numbers on start issue page.

Empty selected page Numbers: Do a multiple selection of pages and empty page numbers.

Analyze as List: If the recognition of a part in a document that looks similar to list lead to a bad result (for example too many zones), the special feature Analyze as List might provide a much better result. To use this feature, select the zone or zones belonging to the list. Then, move the cursor to the tree view (usually on the left side of the user interface) and do a right hand click on one of the highlighted entries. A context menu appears. Choose 'Actions' and here 'Analyze as TOC'. docWizz now reanalyzes the zone or zones and treats it as an object as a list. The result should be much better than the one before.

Move Page Up: Page is moved one page up.

Move Page Down: Page is moved one page down.

Reset OCR: Resets OCR of all selected blocks.

Set OCR to document settings: Set the OCR engine and version, the language and the font type of the selected zone(s) the same as the document's. The action can be used to revert all the OCR changes made on a zone.

Set OCR engine: Open a list with the supported OCR engines, from which the user can select on OCR engine to perform the OCR on selected zone(s).

Set OCR Language: Select language from the list appearing.

Set OCR Type: Select from list Antiqua, Fraktur, Typewriter, Fraktur+Antiqua or Typewriter+Antiqua.

Set OCR reading type: Select from list Auto, Horizontal stripes (by line), Vertical stripes (by column).

Set HTRID: Select from list the available models and HTRID values. The menu entry "Set HTRID" is displayed if the selected zone has OCR Engine Transkribus. For Transkribus, since it deals with a lot of different writing styles, both handwritten or printed, language is less important, more important is the writing style, or maybe custom model for a particular collection. This is why you need to set also the ID of the model used.

Reset error / Set error

Resets or sets the error status of the current item. Especially used for page number sequence checking, pages where the page numbers are out of sequence get a flag "Error". An operator may accept (reset, for example when indeed there is a new sequence started) or they may set again the error status in case he detects something wrong by his own.

Refresh Item

If changes are not shown immediately, press Refresh to update the item.

Properties

Selecting Properties would open the Properties dialog that shows information about the current text block.

Navigation

It is possible to concentrate on the erroneous items in the page sequence. To jump easily from one suspicious page to the next simply use the red arrow buttons (up/down) on the left toolbar.

Extended tool bar for page numbers



The toolbar works on selection of pages in Review

On multiple selection buttons work as follows:

- | | |
|--------|---|
| PN | Is displayed and works on single and multiple selection, on single selection it fills the page number series from current selection to the last page, on multiple selection (> 2 pages selection) fills the selected series of pages. |
| PNCopy | Is displayed and works on single and multiple selection, on single selection it fills the page number series from current selection to the last page, on multiple selection (> 2 pages selection) fills the selected series of pages. |
| PN(/) | Is displayed and works on single and multiple selection, on single selection it fills the page number series from current selection to the last page, on multiple selection (> 2 pages selection) fills the selected series of pages. |

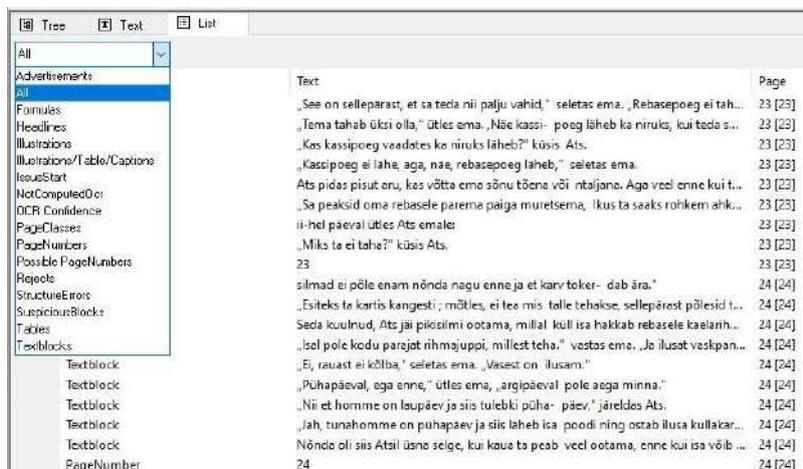
<p>Page Sequence tree view (as a review it will work also on single selection in tree).</p> <p>The significance of the buttons is related to types of filling page number series, they are also explained in the description area in the status bar.</p>	<p>+PN(/)</p> <p>+2PN</p> <p>+PN(j)</p> <p>8PN</p> <p>16PN</p> <p>Empty</p>	<p>Is displayed and works on single and multiple selection, on single selection empties the page number series from current selection to the last page, on multiple selection (> 2 pages selection) empties the selected series of pages.</p> <p>Is displayed and works on single and multiple selection, on single selection it fills the page number series from current selection to the last page, on multiple selection (> 2 pages selection) fills the selected series of pages.</p> <p>Is displayed and works on single and multiple selection, on single selection it fills the page number series from current selection to the last page, on multiple selection (> 2 pages selection) fills the selected series of pages.</p> <p>Is displayed and works on minimum 8 pages selection.</p> <p>Is displayed and works on minimum 16 pages selection.</p> <p>Is displayed on single and multiple selection and empties the page number only on selection.</p>
--	---	---

3.4.4.2.1 List view in Review page sequence task

The **List view** shows only specific zones such as headlines, authors or footnotes to perform a powerful correction without checking each page. Swapping through the headlines, for example, shows only those pages where headlines have been identified automatically.

To choose the type of zones that should be verified, choose the desired one in the related drop down menu. Alternatively, you can use shortcuts

('h' for headline, 'a' for author, 'f' for footnote, 't' for table or text block (press 't' twice)).



Select entry from drop-down menu:

- Advertisement** To verify the indicated advertisements.
- All** Show all object classes.
- Illustrations / Table / Captions** To verify the indicated illustrations, tables or captions.

Headlines	Swapping through the headlines shows only those pages where headlines have been identified automatically. The image view on the right always shows the image page with recognized zones of the related item. You're always able to merge two or more zones in the right part of the screen, for example if one headline has been split into two or more zones.
IssueStart	Especially for serials - used to check how many issues you are creating and if the issues contain the correct amount of pages. Purpose: having fast verification of tagged issue starts easy double checked in a bunch of hundreds of pages
Not computed OCR	Show all classes where no OCR is done. Loads high resolution image from server and tries to compute OCR.
OCR Confidence	To verify the indicated OCR confidence.
Page classes	Show page class "imagePage".
Page Numbers	To verify the indicated page numbers.
Possible Page Numbers	
Rejects	Show list of all rejects. Contains the reason why the reject was raised. User action tells witch rejects has been accepted or not. Toggle status: Rejected - Accepted.
Structure errors	
Suspicious blocks	To identify and correct or delete untypical elements like noise blocks.
Tables	To verify the indicated tables.
Text blocks	You're always able to merge two or more zones in the right part of the screen, for example if one headline has been split into two or more zones.

3.4.5 Structure

Working in the structure step is different for monographs or newspapers.

3.4.5.1 How to use the Review issues task

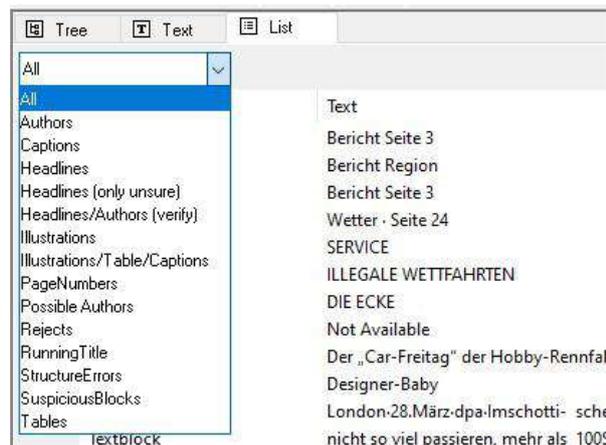
This task allows you to check if the main parts of the document are separated well or not. Books that have been processed as **Monographs** are separated into three sections: Front, Main and Back. You may first check, if each section exactly contains all those pages, which belong to it.

Tip: You can use the **Fast correction of zones** feature.

In **Review issues**, docWizz separates the document into basic parts: Issues for newspapers or to Front, Main and Back for monographs.

How to use lists in Review issues task

In Review issues task, the list view look like this:



The **List View** shows only specific zones such as headlines, authors or footnotes to perform a powerful correction without checking each page. Swapping through the headlines, for example, shows only those pages where headlines have been identified automatically.

To choose the type of zones that should be verified, choose the desired one in the related drop down menu. Alternatively, you can use shortcuts

('h' for headline, 'a' for author, 'f' for footnote, 't' for table or text block (press 't' twice)).

Authors

To verify items classified as author, choose this entry in the drop down menu. You're always able to merge two or more zones in the right part of the screen, for example if one headline has been split into two or more zones.

All

Show all object classes.

Captions

To verify the indicated captions.

Headlines

Swapping through the headlines shows only those pages where headlines have been identified automatically. The image view on the right always shows the image page with recognized zones of the related item. You're always able to merge two or more zones in the right part of the screen, for example if one headline has been split into two or more zones.

Headlines/Authors (verify)

Calling Headlines/Authors helps to verify the correct recognition of the items headline and author. As these items might look quite similar, it is important to check them. To change a type of an item, just right click on the related item in the List View and double click at the desired zone type. You're always able to merge two or more zones in the right part of the screen, for example if one headline has been split into two or more zones..

Headlines (only unsure)

Calling Headlines (only unsure) helps to verify the correct recognition headlines. Sometimes simple text blocks look like headlines or headlines look like text blocks. Whenever dW indicates a probability that an item might be a different zone type, it shows it in this view. To change a type of an item, just right click on the related item in the List View and double click at the desired zone type. You're always able to merge two or more zones in the right part of the screen, for example if one headline has been split into two or more zones.

Illustrations

To verify the illustrations.

Illustrations, Tables, Captions

To verify illustrations and tables and captions.

Page Numbers

To verify the indicated page numbers.

Possible Authors

To verify the indicated possible authors.

Rejects

Show list of all rejects. Contains the reason why the reject was raised. User action tells witch rejects has been accepted or not.

Toggle status: Rejected - Accepted.

Running title

To verify the indicated running titles.

Structure errors

Suspicious blocks

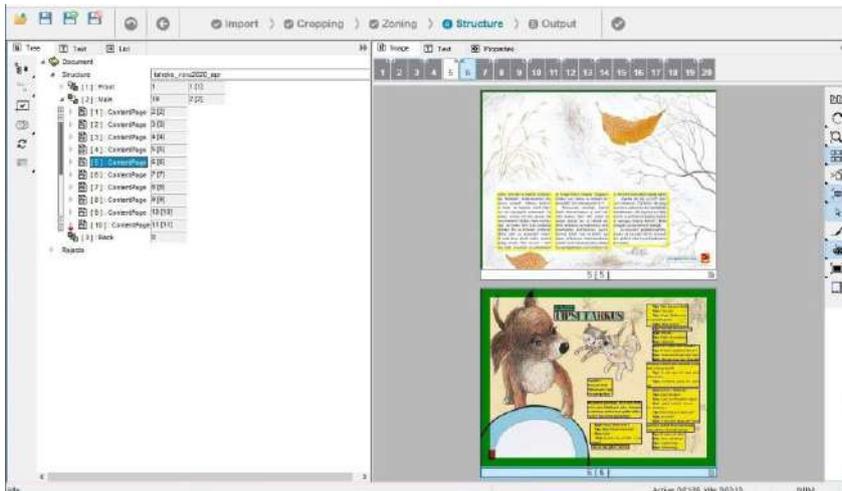
To identify and correct or delete untypical elements like noise blocks.

Tables

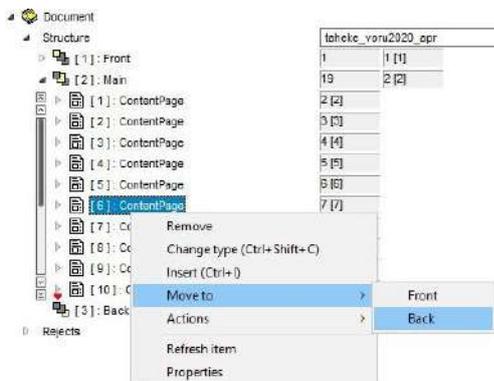
To verify the indicated tables.

3.4.5.1.1 Structuring monographs

The Front part contains title page, table of content etc. The Main part contains the chapters or, if a serial is processed, the issues. Last but not least, the Back part may contain an appendix. This is also an essential task in order receive the correct structure in the end. To ensure this, the verification of the recognized headlines is the most important task. This can be done choosing the 'list view'. In this view the user is able to check and maybe correct just the list of headlines, or unsure (unsure in terms of possible errors) headlines.



In case you find some misarranged pages you can easily move them to the section, they belong to.

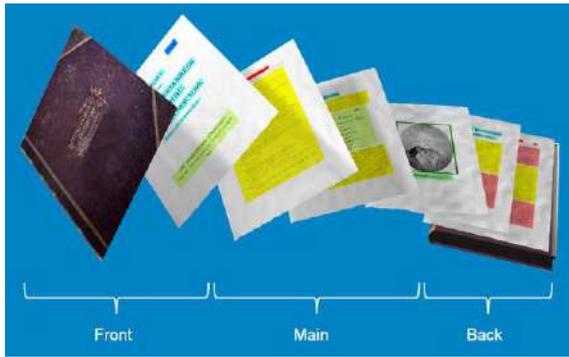


To move pages to another section put the cursor on the icon of the misarranged page in the tree view and press the right mouse button.

This moves the current page along with the previous/following pages to the Front/Main or Back part of the current issue.

Front-Main-Back

If a monograph is processed, you just deal with the correlation of pages to Front, Main and Back.



docWizz separates documents into three major parts: Front, Main and Back. Front consists of elements (pages) such as title page, preface, table of content, etc. All the content (chapters or, if a serial is processed, issues). Back may contain for example Appendixes. At this stage, only a determination into Front, Main and Back is done. The more detailed distinction takes place in the next task within docWizz.

In order to easily change the results of automatic recognition, you can move a sequence of pages to a different part.

Front-Issues-Back

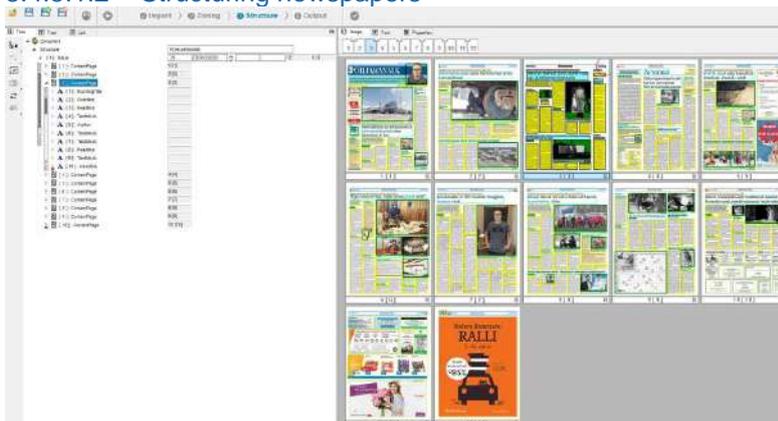
If the current document has been processed as Serial you must check if all Issues have been detected and separated correctly. You may also verify the front and back part of the entire book. Pressing the right mouse button on the issue icon in the tree view may insert or remove issues. If you delete an issue, its content will be moved and assorted to the previous issue.

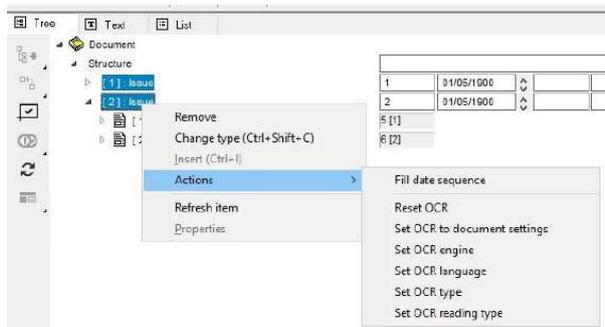
If a serial is processed, the user allocates the pages to Front, Issues and Back. Issues contain Front, Main and Back levels, too. So the determination is more complicated when working with serials.

Note: It is not possible to move multiple pages at once. The sequence shall remain unchanged. So you just select the last/first page to be moved. All pages before/after are automatically moved to the different section.

See [Tree view](#) chapter for explanations of the other entries from the context menu in detail.

3.4.5.1.2 Structuring newspapers





See [Tree view](#) chapter for explanations of the other entries from the context menu in detail.

Insert empty Issue

To insert an empty issue, do a right-click in tree view on the imagePage where the issue shall begin. Choose 'Insert'. docWizz automatically inserts the hierarchy structure in the tree, containing the new issue with the subentries front, main and back.

To build a further issue from an existing one, do a right click on the related page in the tree view and select 'Insert'. The current issue will be split beginning with the selected page. The new issue that is built contains the content from that page on up to the end of the current issue.

Delete Issue

Deletes the hierarchy information about these issue and moves the content in the one above. The first issue can't be deleted.

To delete an issue, do a right-click on an issue entry in the tree view. Choose 'Delete'.

Move to Issue

Right-click on a single selection consisting of Image-Page in Outline Tree View and select æMove to Main/Issue.

If the current Image-Page belongs to a Front section then the action moves the current Image-Page along with all the next Front pages to the Main/Issue of the current selection. If the current Image-Page belongs to a Back section then the action moves the current Image-Page along with all the previous Back pages to the Main/Issue of the current selection.

Restore Page Sequence

Right-click on an Image-Page in the Tree View and select Actions/Restore Page Sequence.

On any page that is applied Restore Page Sequence acts as an undo operation to previous æMove to Front of Volume or Move to Back of Volume thus restoring the initial relative position of pages.

Auto detection of dates

Date detection, is automatically done in the automated Recognize issues step (done before Review issues) and is available for newspapers and serials configurations. The feature works without any scripts or custom actions.

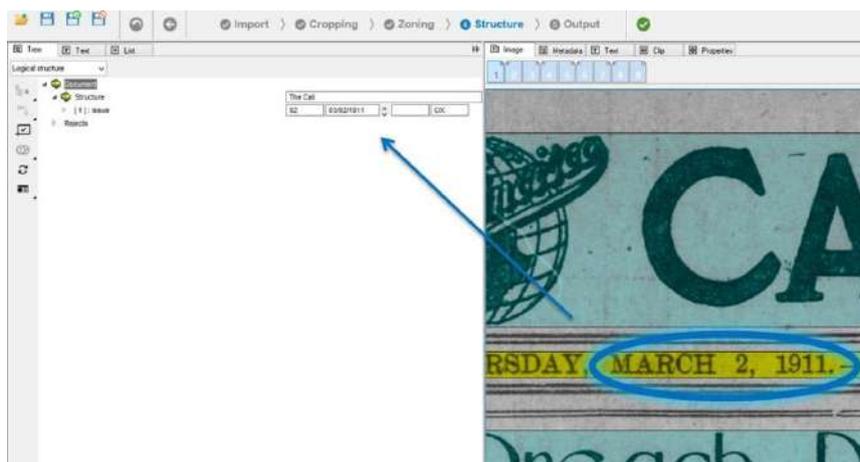
In the Recognize issues step, the OCR from RunningTitle zones, that are placed in the top part of the image from an issue, are checked and compared with a "date-like" format string. If one date is found in more that x% of the RunningTitles, that date is set as the IssueDate. Otherwise, the IssueDate will be left empty.

The number of occurrences depends on the number of pages in that issue:

- for 6 pages or less - at least 2 identical dates

- for 7-10 pages - it's number of pages * 0.3 (ex: for 10 pages, 3 identical dates)
- more than 10 pages - number of pages * 0.2 (ex: for 20 pages, 4 identical dates).

Only the RunningTitles from the top part of the image are used!



3.4.5.2 How to use the Review structure and text task

The last task in interactive processing is the task **Review Structure and Text**. In this task you can perform all corrections you might want to do. This includes classification of blocks (e.g. footnotes, pictures, picture texts), hierarchy (rearrange chapters, subchapters, contributions), move items to front and back etc.

3.4.5.2.1 Subtask: Review structure

In this task each layout element must be classified correctly by its type. This is especially important for **headlines** as the logical document hierarchy is based on these results.

An operator might have already specified some elements earlier in Review zoning. This means that only Headlines must be checked.

How to work with the List view

To do so, change to **List view** and choose **Headlines** in the drop down menu below.

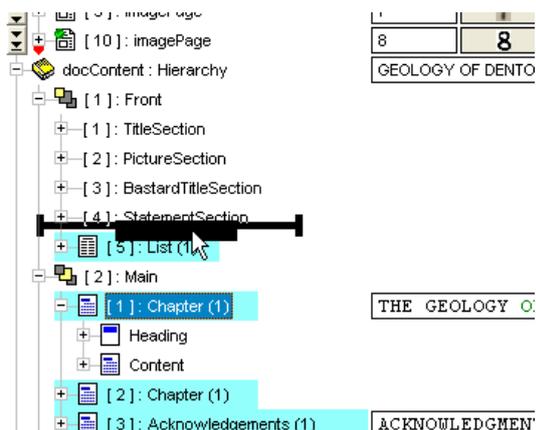
Entity type	Text	Page
Headline	R.F. PORT OF DEPUTATION	1 [1]
Headline	REPORT OF DEPUTATION.	3 [1]
Headline	APPENDIX A.	22 [20]
Headline	AUSTRALASIAN WESLEYAN METHO...	22 [20]
Headline	INSTRUCTIONS.	22 [20]
Headline	APPENDIX B.	23 [21]
Headline	THE ORGANIZATION OF FIJI CIRCU...	23 [21]
Headline	APPENDIX C.	27 [25]
Headline	FRIENDLY ISLANDS.	27 [25]
Headline	APPENDIX D.	30 [28]
Headline	*REPORT OF THE DEPUTATION TO F...	33 [31]

- For changing a headline to normal text right click on the element and choose **Change Type**. You can also use the shortcut (T).
- Check one headline after another by scrolling down with the cursor.
- Having finished press the green process button on top right.

Note: Illustrations are not allowed in Heading.

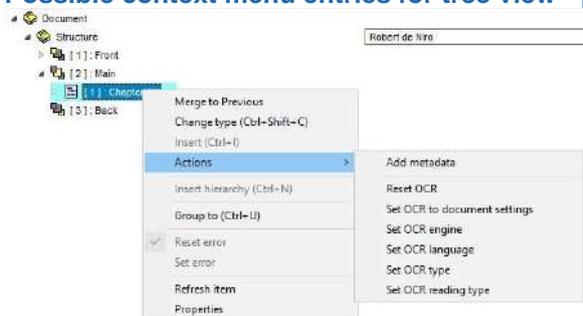
How to work with the Tree view

Sorting by drag & drop



Select e.g. a chapter from the Front section and drag it to the Main section. Wait until the black line appears and release mouse at the desired place.

Possible context menu entries for tree view - pages structure:



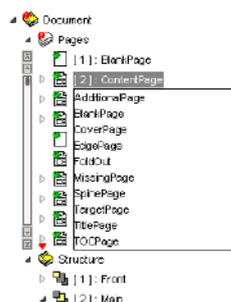
Merge to previous

Only available for Document: Hierarchy, Article level.

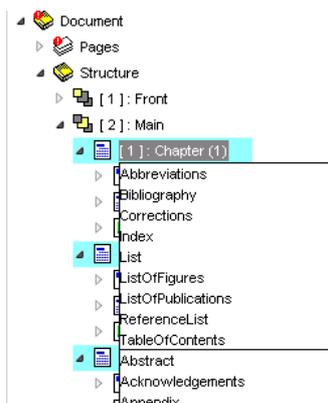
A separate article with own headline will be merged under the previous article's headline.

Change Type

You can use as well the tree view to assign a zone to another zone type. Pressing the right mouse button, with the cursor on top of a text block entry, opens a context menu. Selecting **Change Type** opens a small list window attached to the text block symbol. Select the desired type name (Advertisement, Formula, Illustration, Table or Vertical Textblock) in the list and assign it with a double click of the left mouse button to the zone.



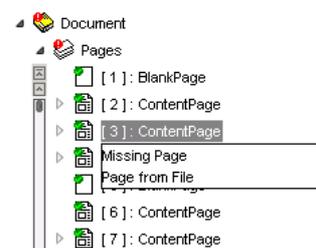
In Document - Structure: You can change articles in chapter or section and vice versa.



Insert

After the system has calculated the new entry and has integrated it into the context of the document, the new zone is highlighted in yellow and is defined as text zone. You may assign the zone to another zone type as described above.

Select **Insert** in the context dialog. Choose between **Missing Page** and **Page from File**.



After the system has calculated the new entry and has integrated it into the context of the document, the new zone is highlighted in yellow and is defined as text zone. You may assign the zone to another zone type as described above.

Insert hierarchy (Heading)

In terms of hierarchy levels, you can move any entry in the tree (for example chapter or subchapter), including the subentries, to a higher or lower level.

To make a subchapter into a chapter, select the subchapter's entry and right click on that highlighted entry in the tree. Then, choose Actions-**Level up**.

Right-Click on a Heading in the Tree view and execute **Insert Hierarchy**. **Retrieves** all the blocks from Heading after the current block (and including it) plus all the Content and creates a new chapter in the contents. The newly created chapter will contain the previous-mentioned entities, on the same level as the Heading that the command was applied onto. As a secondary effect, the first block in the succession is converted into a Headline.

"Level Up" / "Level Down" (Chapters in text section)

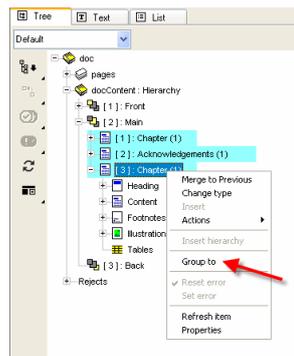
Select one or multiple chapters in Text-Section-like entities and execute **Level Up** or **Level Down**. **Level Up / Level Down** chapters in Text-Section-like behaves almost the same as **Level Up / Level Down** commands in the other chapters with one exception: if they are to be found on the first level (Subtext section) the **Level Up** extract the Text-Section and creates a new Text-Section with it and all that it is contained below it. For the moment this command works only for a single selection. The following properties of the chapter are rebuilt: Illustration-list, Tables and Text Notes.

"Level Up" / "Level Down" (Chapters, sublists and entries)

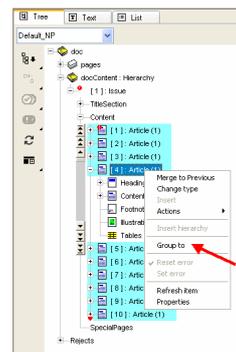
Select one or multiple chapters, sub-lists and entries and execute **Level Up** or **Level Down**. Increases or decreases the position of the selection in the hierarchy.

Group to

Monographs



Newspapers



Group to (other items): Right-Click on multiple selections consisting of Paragraphs, headlines, Text-Blocks, Authors, Poems and select “Group To”. This will group them into a Chapter-like entity. You can select the Chapter type from a list box. In the newly resulted chapter the first block is converted to Headline and all the Paragraphs are converted to Text-Blocks.

Group To (Page-like entity): Right-Click on a multiple selection consisting of Page-like entities and select “Group To”. You can group them to a List-like entity or Text-Section-like entity (Appendix, Bibliography, Dedication, Introduction, Necrology, Preface).

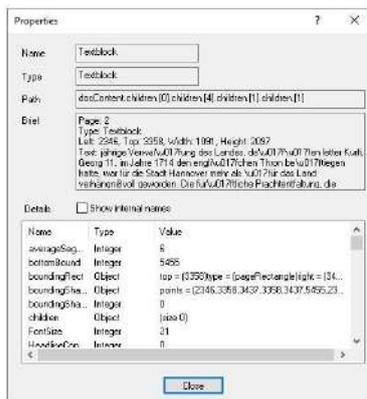
Group To (Sub-list and entries): Right-Click on multiple selections consisting of Sub-Lists and/or Entries and select “Group To”. You can group them to another Sub-List and/or entry.

Refresh Item

If changes are not shown immediately, press **Refresh** to update the item.

Properties

Selecting **Properties** would open the **Properties** dialog that shows information about the current text block.



Entries under **Actions** depend from where the context menu is opened. So not all of the following described actions are available at any time.

Add metadata

Adds metadata of the selected image page.

Make same level

Only available for docContent: Hierarchy. Moves subentries (sub articles) to the level above.

Sort blocks in article

Only available for docContent: Hierarchy.

Sort TOC items

Only available in docContent: Hierarchy and the List entry.

Fill page number sequence

Use, if you have to fill some page numbers in a row. Therefore, select the last page that contains a page number in the middle column and select all following pages, which do not contain page numbers in the middle column, too. To do the multiple selection, hold the (Ctrl) key while clicking on each page. Then, do a right click on any highlighted page item in the tree view and choose Actions-Fill Page Number Series in the context menu.

Note: It is possible to avoid the renumbering of page numbers even if a new issue was tagged in the serials configuration. This is due to that serials often times have continuous page numbers over the whole year and the operator has many double work to do to change back the page number series. Therefore auto-renumbering for Serial while setting issue start is disabled in the default configuration.

Empty page number

Do a multiple selection of pages and empty page numbers.

Reset OCR

Resets the OCR.

Set OCR to document settings

Set the OCR engine and version, the language and the font type of the selected zone(s) the same as the document's. The action can be used to revert all the OCR changes made on a zone.

Set OCR engine

Open a list with the supported OCR engines, from which the user can select on OCR engine to perform the OCR on selected zone(s).

Set HTRID

Select from list the available models and HTRID values. The menu entry "Set HTRID" is displayed if the selected zone has OCR Engine Transkribus. For Transkribus, since it deals with a lot of different writing styles, both handwritten or printed, language is less important, more important is the writing style, or maybe custom model for a particular collection. This is why you need to set also the ID of the model used.

Set OCR Language

Select language from the list appearing.

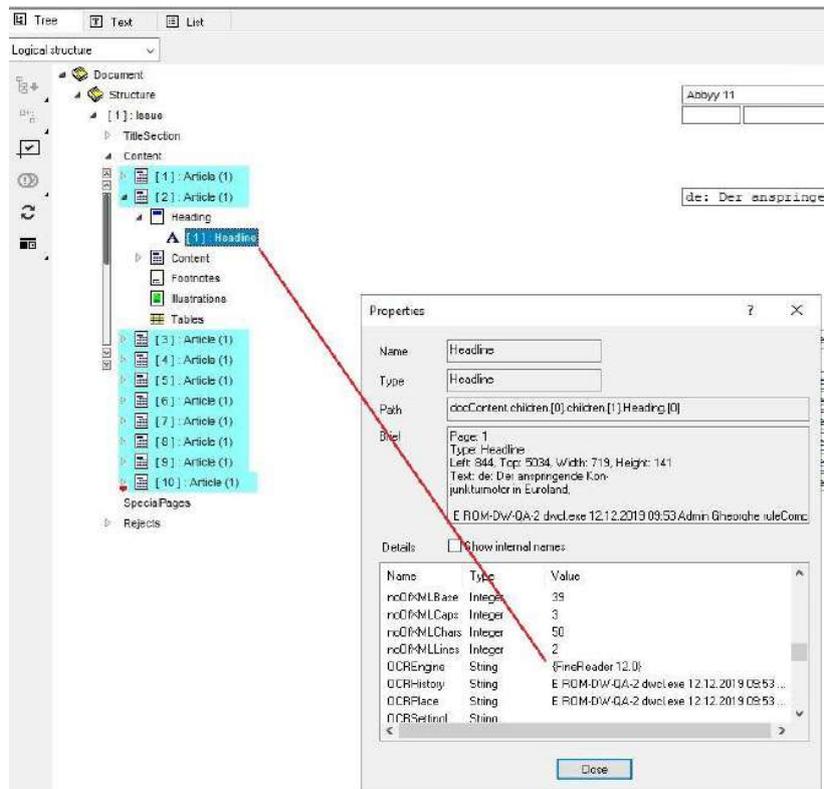
Set OCR Type

Select from list Antiqua, Fraktur, Typewriter, Fraktur+Antiqua or Typewriter+Antiqua.

Set OCR reading type

Select from list Auto, Horizontal stripes (by line), Vertical stripes (by column).

A right click to open the in context menu. Select Properties shows, among others, the OCR license used.



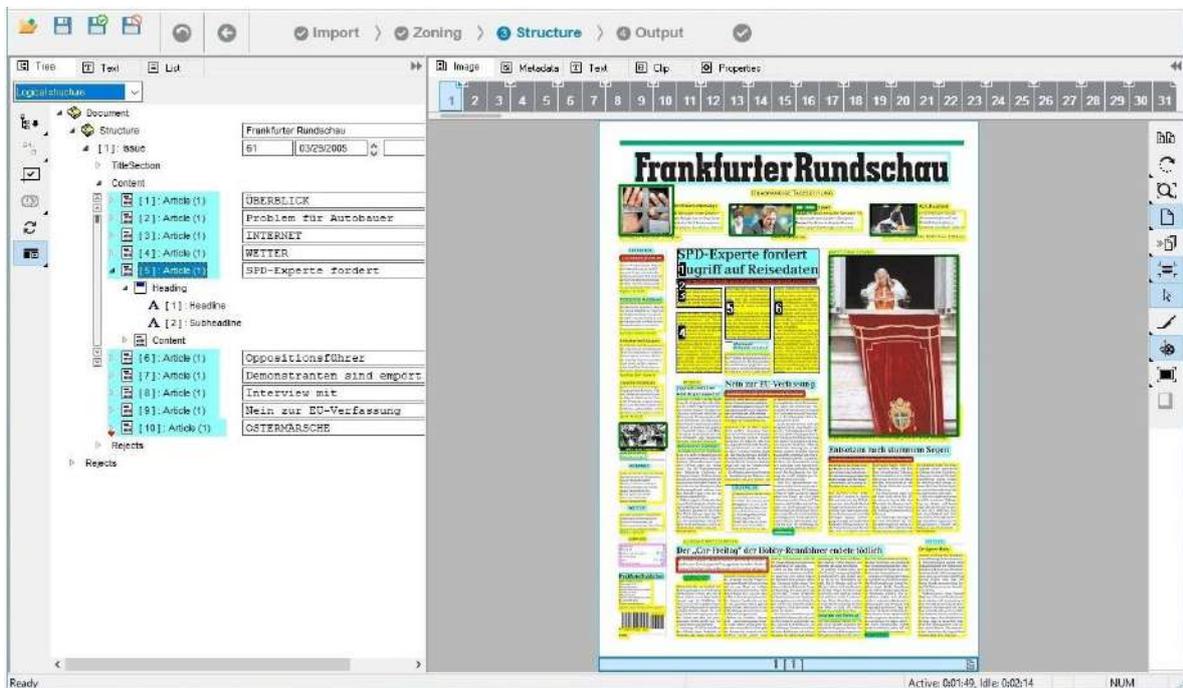
Group mode

The  **Group mode** button in Review structure and text task toggles grouping mode. You select a structure element (e.g. an article) in tree view and then you simply click on those zones (or drag a frame to cover more than one) that should belong to the article as well but are not contained yet. The cursor turns to .

The sorting of article zones takes into account if text is written left-to-right or right-to-left. If the document have set only a single language as document language, then the reading order specific to that language will be used. In case multiple languages are set to document the decision regarding sorting will be taken based on the languages of individual text zones, part of the article, if OCR text is available. The majority of the zones will give the sorting algorithm (f.e. if an article contains 5 zones with Arabic text and a single one with English text, then right-to-left sorting is used, specific to Arabic language). This rule apply on all sorting actions in docWizz. You can overwrite the sorting by TCL scripting, adding into your project a TCL script named `[projectcfg]-SortBlocksInArticle.tcl` implementing custom sorting algorithm.

Having the content of the article selected and clicking on every text block will create a paragraph for each zone if it is not created automatically by detection. This is useful for article blocks ordering.

- To add a missing zone (e.g paragraph) to an article select a structure element (e.g. an article) in tree view on left hand side.
- Click the group mode  button.
- The cursor turns to . Click on the missing paragraph(s) in image view on right hand side or drag a frame to cover more than one.



Click in the reading sequence so that the paragraphs will be in the correct order (small numbers in the edge of each zone.)



The paragraph will be added automatically to the article in the tree view.

Functionality works when selecting almost any entity in Review structure and text tree:

- Grouping entities to TitleSection
- Grouping entities to IllustrationStruct\TableStruct
- Grouping entities to Heading
- Grouping entities to Paragraph
- Ordering zones in the article content using Grouping Mode

The selected entities are filtered before being added, only the valid entities can be added to the selected tree structure. For example a headline cannot be added to an IllustrationStruct or a Paragraph. On the other hand a textblock can be added to the heading of the article or TitleSection of the document.

Grouping mode can be used on structures inside the Article/Section/Chapter. For each structure, click and drag adds only specific zone types.

E.g.:

- For Heading structure, with click and drag you can add Headline, Subheadline, Overline, Author and Textblock.

- For Content structure you can add: Textblocks (which are grouped into Paragraphs), Formula and Author (you can also use grouping mode on Paragraphs to add new Textblocks to this structure).
- In Footnotes structure, only Footnote blocks can be added.
- **Illustrations** – only Illustrations and Captions can be added to this structure. If both the illustration and its caption are selected, they are inserted in the same IllustrationStruct. If the caption is selected first and then the Illustration, the caption is inserted in one IllustrationStruct (an existing one that contains an illustration, and if it doesn't exist, it creates one) and the illustration is inserted in different (new) IllustrationStruct. If the illustration is selected first and then the caption, the illustration is inserted in a new IllustrationStruct and the caption search for the best matching IllustrationStruct (you can use grouping mode on IllustrationStruct to add new Illustrations or Captions to this structure).

Sorting

Sorting can be done on every structure of the docContent: Hierarchy. Sorting can be made using click on the blocks of the structure or using click and drag – only for empty structures.

Sorting by a click

On a structure that contains elements, you can sort the elements using click.



If you have the content of an article with the elements in a random order (like above – 1, 4, 6, 5, 3, 2), you can click on each block in the order you would like to have. For sorting the article above, you will have to click on the blocks in this order: 4 -> 6 -> 5 -> 3 -> 2. Each block clicked on becomes the last one, so when pressing block 4 in the example, the new order will be: 1, 6, 5, 4, 3, 2.

You can get the default order in the article by selecting the Article structure and clicking on one of its elements (the default order is: from top to bottom and from left to right).

The same process can be used to sort blocks in all structures (Heading, Content), except IllustrationStruct and Tables. In both structures the order of the blocks is determined by the order they are added to this structure.

E.g.: You have 3 illustrations and only one caption in this order:

The screenshot shows a document editor interface. On the left, a 'docContent: Hierarchy' tree lists the document's structure. Under 'Content', there are four 'Article' blocks (1-4) and ten 'Paragraph' blocks (1-10). Under 'Illustrations', there is one 'IllustrationStruct' containing four 'Illustration' blocks (1-4). The main window displays a page layout with four numbered elements: 1. A cartoon character, 2. A photograph of people holding flags, 3. A photograph of a man in a hat standing in front of a red wall with white patterns, and 4. A text block with a caption below it. The caption text reads: 'Sarajevo har bygget siden 1978. Olympiske vinterleker krever mellom 1,5 og 2 milliarder kroner i investering. TV, coca cola og amerikansk øl betaler det meste. Nå er alt klart til å prøve arenaer som vi nordmenn er sterkt interessert i. Men Bosnias hovedstad er mest opptatt av hva som skal skje etter lekene i 1984.'

You can't sort the blocks from this IllustrationStruct using click. If you select Illustrations structure, and click on each block from that structure, each Illustration will be grouped in one IllustrationStruct and the caption linked to the nearest Illustration.

This screenshot shows the same document editor, but with the 'Illustrations' structure selected in the hierarchy. The page layout has changed: the cartoon (1) is at the top, followed by the photo of people with flags (2), and then the photo of the man in a hat (4). The text block (3) is now positioned directly below the photo of the man in a hat. The caption text is the same as in the previous screenshot.

You can get the default order in the IllustrationStruct by selecting the Article structure and clicking on one of its elements.

Sorting using click and drag

You can choose the order of the blocks in a structure using click and drag, but only on empty structures or on structures that you want to add something at the end of it.

E.g.: You have to add the Textblocks in the Content of the article from right to left.

The screenshot shows a content management system interface. On the left, there is a 'Structure' panel with a tree view containing: Article (1), Article (1), Article (1), FRA RUSSL, Heading, Content, Footnotes, Illustration, Tables, and SpecialPages. Below this is a 'Reverts' button. The main area displays a preview of a page titled 'FRA RUSLAND IDAG' with the subtitle 'DE KOMMUNISTISKE TEORIERS FALLITT'. The preview shows a newspaper-style layout with multiple columns of text and a large image. Red arrows point from the right side of the preview towards the 'Content' section of the structure, indicating the process of dragging content blocks from the right side of the preview into the 'Content' section of the structure.

You can start clicking and dragging the columns one by one from right to left. The first column added will be placed last in Content structure, the second one will be placed second to last and so on.

This screenshot is similar to the one above, but with red boxes highlighting specific text blocks in the preview. Red arrows point from these highlighted blocks to the 'Content' section of the structure, illustrating the 'click and drag' process. The highlighted blocks are located in the rightmost column of the preview, and the arrows indicate they are being moved into the 'Content' section of the structure.

If you want to get the default order in the content, select Article structure and click on an element. Click and drag sorting can be used on all structures: TitleSection, Heading, Content, IllustrationStruct etc.

Title section

On TitleSection structure you can use click and drag to add blocks in the structure:

Headline, Subheadline, Overline, Textblock, Illustration (creates an IllustrationStruct inside the TitleSection),

Illustration + Caption (creates an IllustrationStruct inside the TitleSection and adds both elements in it.

Only the Caption can be added in TitleSection if the Caption is linked with an Illustration in an IllustrationStruct. If the Caption is not linked, it is not added to TitleSection.

Other zones like Advertisement, Table, RunningTitle, PageNumber can't be added using click and drag or dragging using (Ctrl).

Drag over the whole TitleSection area of the page (including advertisements):

The screenshot shows a document editor interface. On the left, a tree view shows the document structure: 'doc' contains 'pages', which contains 'docContent: Hierarchy' (Blond) and '[1]: Issue' (1, 12/10/2011). Under '[1]: Issue', there is a 'TitleSection' containing 'Content' and 'SpecialPages'. A table below the tree lists five issues: Issue 2 (12/11/2011), Issue 3 (12/12/2011), Issue 4 (12/13/2011), and Issue 5 (12/14/2011). The main content area displays a newspaper page with a masthead for 'American Register' and 'The Morning News'. The page is filled with various advertisements, including 'AU BON MARCHÉ', 'GRAND HOTEL ROME', 'TIFFANY & CO.', 'JEWELRY MARTEL JEWELRY', and 'ROMA HOTEL DU QUIRINAL'. The masthead area is highlighted with a blue border, and the table of contents is visible on the left side of the page.

This image shows what is added in TitleSection:

This screenshot shows the same document editor interface as the previous one, but the 'TitleSection' is expanded to show a list of elements: '[1]: IllustrationStruct', '[2]: Headline', '[3]: IllustrationStruct', '[4]: IllustrationStruct', '[5]: Textblock', '[6]: Textblock', and '[7]: Textblock'. The main content area shows the same newspaper page, but the masthead area (including 'American Register' and 'The Morning News') is now marked with red 'X' symbols, indicating it is orphaned. The table of contents on the left remains the same.

Orphan zones

An autofix is available in ReviewStructureAndText. This adds into structure all orphan zones (that exist on pages and not in hierarchy, and are running elements).

By default this is disabled and could be enabled project based if reject "zones only on page" appear often and a generic solution that try to find best place for the orphan zone is needed, instead on manually fix the issue.

Please ask docWizz support team to activate this feature.

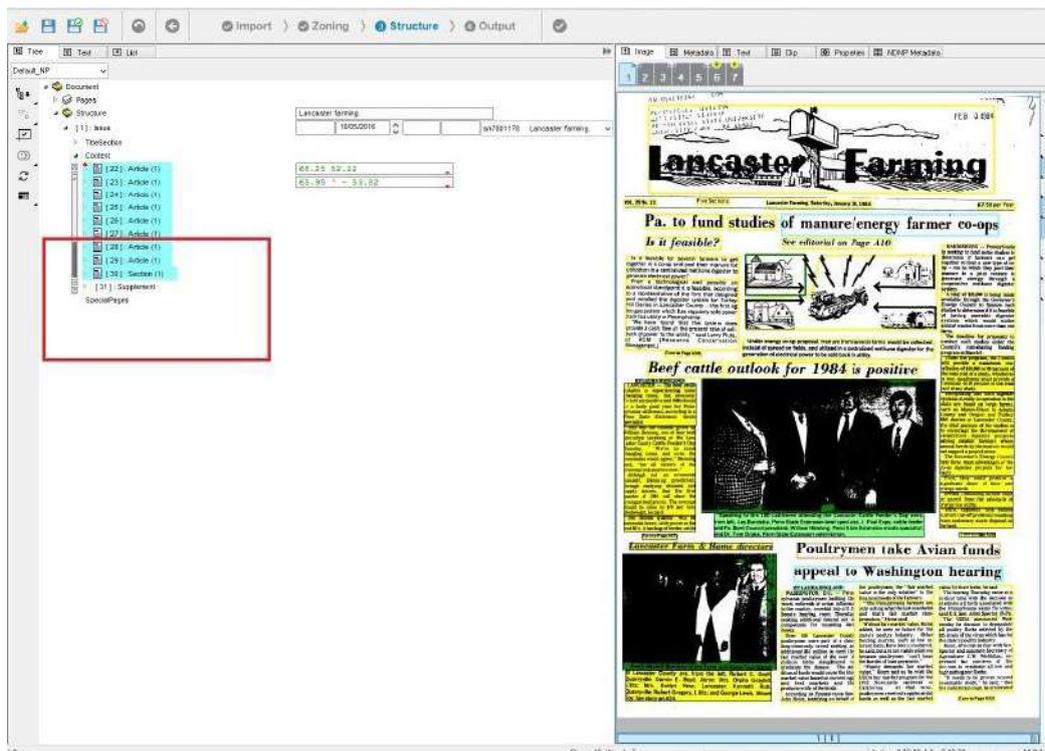
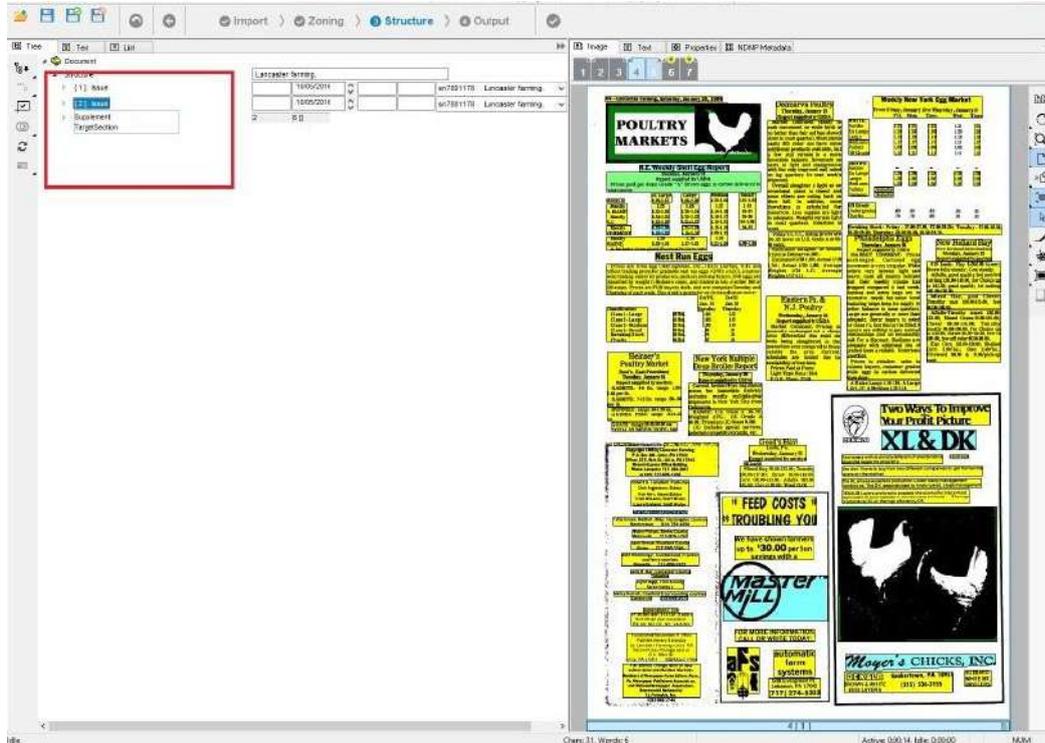
Supplements

After the first page of the supplement is checked in RPS, in the next step RI there will be two issues with the same date.

Here the issue that contains the supplement pages must be changed to "Supplement" using right click on it then change type and selecting from the menu Supplement.

In RSaT step, this supplement will be included in the issue with the same date.

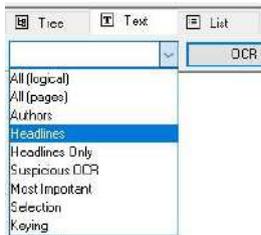
These two screens show how this marking should look:



3.4.5.2.2 Subtask: Review OCR

Text view (left): You may perform text correction here. To do so, select the part you want to correct in the tree view. Open the editor window by the Text tab.

You can select special parts of the document for correction.



Use  and select the document item you wish to correct.

Having selected **Headline**, all headlines are presented in the editor window and can be easily checked and corrected one after the other.



It is very important to use the headline correction in the task before Review Issues. This should ensure to receive good results in terms of structure recognition. The most important issues in the task **Review Structure and Text** are the page classification in the 'Front' as well as in the 'Back' part of the document and the chapter hierarchy in the 'Main' part respectively in the 'Issues'.

The user can change the type of pages in the 'Front' (e.g. changing Title Page into Table of Content) and 'Back' part. You can also correct the hierarchy in the 'Main' part easily by using 'Level up/down'. This moves chapters into a lower (Subchapter) or higher hierarchy level.

Now that you took care of all the structure matters the OCR correction shall be done. There are several tools to simplify the correction at this point. For example, you jump from one suspicious string to the next. Or, you can use the 'Error Word List' to concentrate on the unknown strings.

How to use the OCR text correction



The Text view window for text correction shows at the top the toolbar with icons for the correction.

The top window shows the text, the system has recognized by processing the source images by the OCR-engine. The text is marked with different colors, which is a result of difficulties of special content.

The second window displays the image of the source page so that you are always able to compare the OCR interpretation with the original printing.

The lower window displays the current string to be corrected zoomed in. The string is shown on a yellow background. The corresponding area in the image view is also marked with a yellow background. This makes it easy for you to compare source and text and to correct the text if necessary.

Note: Corrected text – in text view, when correcting a text the user can mark the text as being corrected using (Ctrl) +(D). Starting with ALTO 1.2 there is an attribute for each text line – CS – which is 1 for the corrected lines. Currently in user interface you can only mark a text block as being corrected (not each line), so if a textblock is marked as corrected, in ALTO all its lines have CS=1.

Correction colors

Various colors are used in the Text View.

Color priority:

Color	Description
White	shown if text is inverted
Blue	corrected, spell check corrected
Red	spell check error
Green	numbers or signs
Pink	unsure words
Black or dark green	default

The default colors for the two main text correction modes (with dictionary and without dictionary) are black and dark green.

Black

Example

In many parts of North Texas

Document

Default, Newspaper

Configuration

Without dictionary

Description

To get the color just open a document, and select text view

Explanation

Default color. Click on a word, it then gets "bold". Word changes color from black to blue. Changing to bold identifies that the word has been changed, shown as blue color. Accessing a blue word is meant as verification. If the word is found to be OK by spell check, color changes from blue to black.

Dictionary

For the languages where there is no dictionary, all words are BLACK (except numbers, which are GREEN).

For the languages that have dictionary, there are also RED words (words not found in the dictionary).

Dark green

Example

Paris – France's first

Document

Default, Newspaper

Configuration

With dictionary

Description

To get the color just open a document, and select text view.

Explanation

Default color.

Pink

Example

Joel Resnick ankles AMC high

Document

Default, Newspaper

Configuration

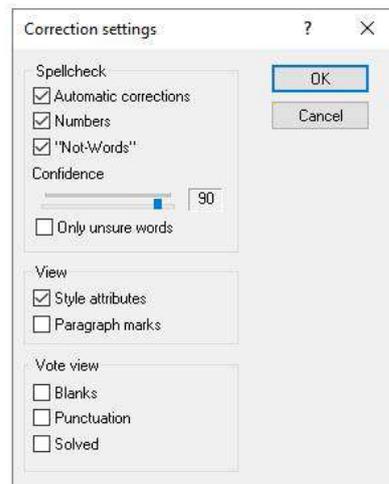
With / without dictionary

Description

The color appears in text view, with "Only Unsure Words" option disabled in OCR settings, but when navigating using the arrow keys over the word, it turns red. Also the confidence bar has to be around 95-100

Explanation

All the words that are considered unsure are displayed in pink.



Open this window by right mouse click on any text in the lower text area. Then select "Settings" from the context menu.

Red

Example

Tarrant County

Document

Default, Newspaper

Configuration

With / without dictionary.

For the languages where there is no dictionary, all words are BLACK (except numbers, which are GREEN). For the languages that have dictionary, the RED words are not found in the dictionary.

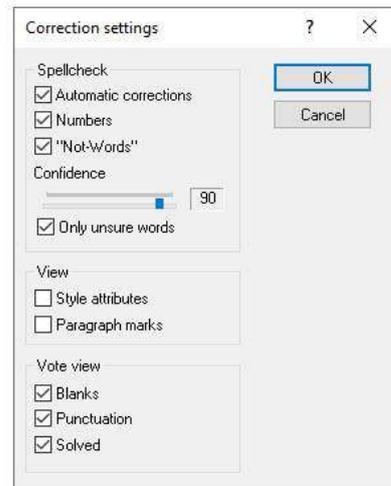
Description

To get the color after selecting text view, navigate using the arrow keys over the word – it turns red (make sure that "Only unsure words" from Correction Settings is not checked).

Explanation

When "Only unsure words" option is not checked the red color is shown if confidence of the word is less than confidence threshold.

When "Only unsure words" option from Correction Settings is checked and the current word has spell check error and the confidence is less than 950 or less than confidence threshold.



Open this window by right mouse click on any text in the lower text area. Then select "Settings" from the context menu.

Light green

Example

1,00 DM D 2955 A
von der IG

Document

Default, Newspaper

Configuration

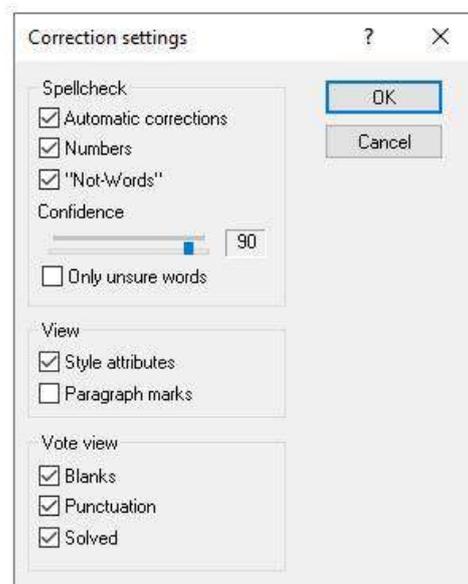
With / without dictionary

Description

The color appears in text view, only on numbers or two letter words, after selecting "Numbers" and "Not-Words" from OCR settings.

Explanation

Numbers and Not-Words are display as light green if both are checked in the setting.



Open this window by right mouse click on any text in the lower text area. Then select "Settings" from the context menu.

Blue

Example

Frankfurter

Document

Default, Newspaper

Configuration

With / without dictionary

Description

When correcting a text in text view, the text turns blue, but after the correction is finished (press 'Space' after the word or click another word) it turns dark green.

Explanation

The color appears on words that are currently being corrected.

Gray

Example

ZEITUNG FÜR DEUTSCHLAND

Document

Default, Newspaper

Configuration

With dictionary

Description

To get the color just open a document and select text view.

Explanation

The color appears only for words that are in dictionary or with different combination in OCR correction settings (uncheck "Tab to not in dictionary").

Highlighted blue

Example

ZEITUNG FÜR DEUTSCHLAND

Document

Default, Newspaper

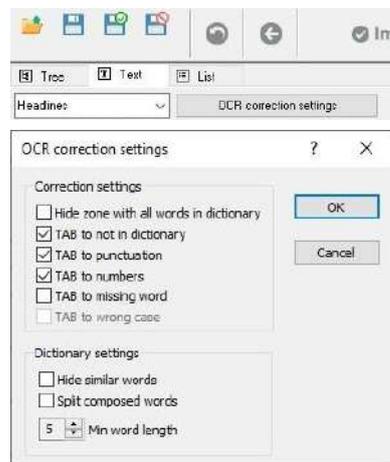
Configuration

With dictionary

Description

Get the color in text view on the left. Select "Tab to missing words" from OCR correction settings button.

Explanation



Correcting errors

To correct an error that is displayed in the text window, you must navigate to the word with the mouse or the cursor keys ← ↑ ↓ → .

The simplest way to do this is to place the mouse pointer on the beginning of the incorrect word and press the (Ins) key to activate the overwrite mode, and then correct the word by overwriting it.

You can also position the mouse pointer behind the last incorrect character of the word, delete the characters one by one with the **Backspace** key and enter the correct characters. The corrected text appears in blue.

When using just keyboard, you walk through the entire text using right and left key.

When cursor is at first position in a line and you press cursor left, cursor shall move to the end of the previous line. When cursor is placed behind the end of a line and you press cursor right, it shall move to begin of the next line.

Pressing (Enter) key when the cursor is in the middle of a text moves cursor position to the beginning of the next row.

Pressing (Delete) key when the cursor is at the end of a text row moves next line of text at the end of current line.

Pressing (Space) key deletes the character after the cursor.

It could be configured to get special characters on OCR correction. Then strike a mapped key quickly to advance to the next mapped character. You will get e.g. the main letter 'e' with variants 'éèë...' or the main letter 'u' with variants='üúû...'. See chapter [OCR](#) in [docWizz ReferenceBook](#).

On all actions that redo OCR, the progress dialog is shown. It tells the current page where OCR runs on. Pressing (Cancel) will stop processing after finishing OCR on current page.

Text correction settings

When you put the cursor into the top text window and click the right mouse button, a context menu is shown. According to your current activities, the functions in this menu are shown as active or inactive and can be used to edit the text.



Cut / copy

Are active, after you have marked parts of the text in the window. You can cut or copy the marked text. In both cases the text will be saved to the clipboard and can be pasted on an other part of the current document or in an other programs to a desired place as often as necessary

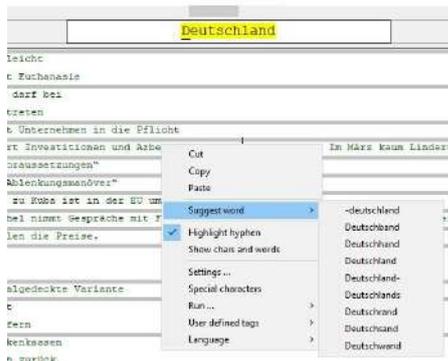
Paste

Is active after you have saved parts of the text to the clipboard, using the **Cut and Copy** functions. You can paste the clipboard texts to a desired place in the document.

Suggest word

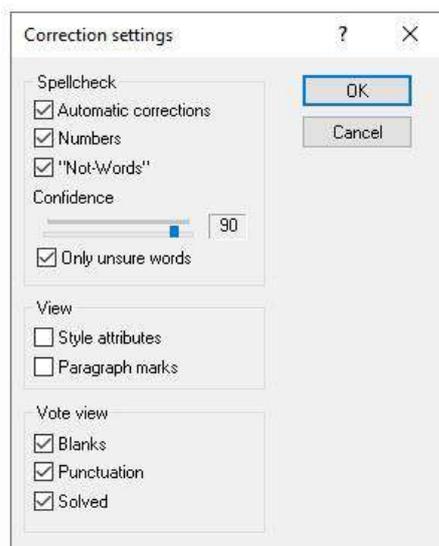
Suggests words, similar to the selected one.

Deutschland



Settings

Calls up the **Correction Settings** dialog box. You can use this dialog to configure the text correction to your special needs.

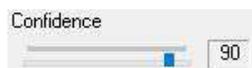


You use the **Spellcheck** area to specify the types of errors that should be submitted to the user for verification. When you select the options provided under **Check**, the system jumps to the corresponding points in the text in order.

If you select **Automatic Corrections** the system jumps to the words that it automatically corrected and which appear in blue.

If you select **Numbers** the system jumps to numbers, which appear in green.

If you select **"Not-words"**, the system jumps to the abbreviations, coded in green.



Using the slider you can choose on the scale a value between 1 and 90 to select the **Confidence** level that the system should use to correct texts.

In the **View** area you can specify the typographical attributes to be displayed in the text window.

Selecting **Style Attributes** means the system will show the text with the original document attributes such as bold text, italics, etc.

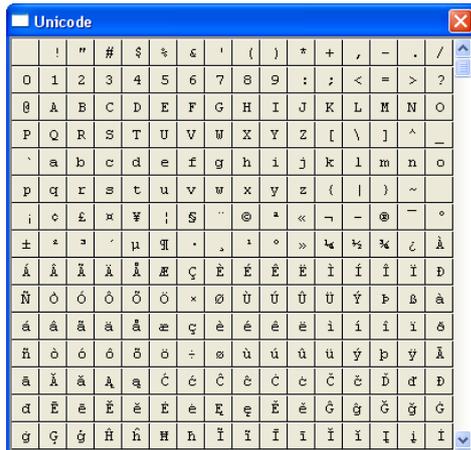
For the left hand text editor this button is disabled. But you can use right click on text - settings - Style attributes checkbox. They can be enabled for Outline Text View.

Selecting **Paragraph Marks** means the system shows the text as it is organized in the original document.

Use the **Vote View** area in OCR correction with dictionary settings.

Special characters

The range of characters used in a printed document may be larger than the character set on your keyboard. However, your PC has more characters than you might suspect. Click the character you want to insert it in your text.

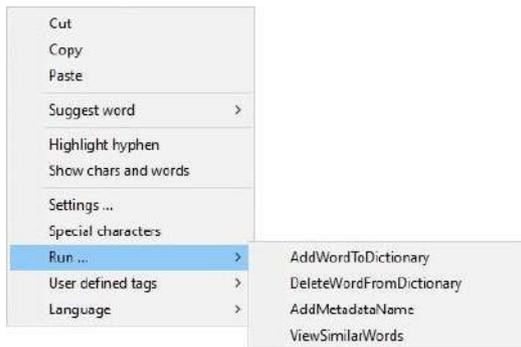


Enter Chinese characters via unicode sequence as with e.g. german special chars. This works with (Ctrl+U 8003)

When Chinese language is set, extra character spacing is used.

Please note: The editor does not support dynamic character width. So when pasting a chinese character into english text, spacing will be incorrect.

Run



Mark word(s) with mouse and open context menu on the word: Run -> **AddWordToDictionary**

```
is and Their Equi-
>R COMANCHEAN
:ation ... 12
```

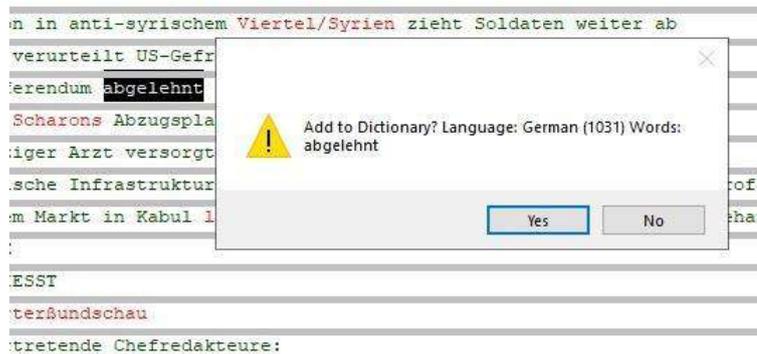
After the word is added to dictionary, it is shown in gray color. One can mark paragraphs, sentences or full text and add it to dictionary.

```
mations and Their Equiv-
EOUS OR COMANCHEAN
xy Formation ... 12
tion 16
```

Words are not separated by space, but are given from code as word object from the line object itself. Because of the min word setting, it does not matter anymore what length have the min words inserted in dictionary.

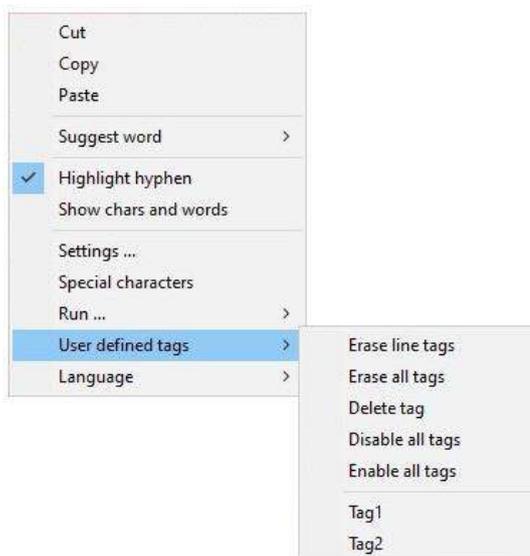
Immediately you can see in the other text fields, that the words are shown in gray, which have been added before.

Selecting an entry here (e.g. AddWordToDictionary) a message will pop up prompting you for confirmation:



DeleteWordFromDictionary deletes word from dictionary.

User defined tags



See [Special feature: Tag](#) chapter on text for details.

Language



Tesseract OCR

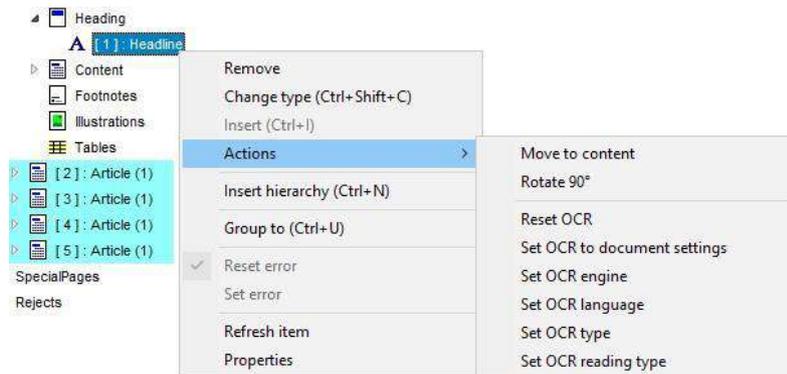
Tesseract OCR is implemented as an alternative to ABBYY Finereader or to support languages not available in Finereader.

Tesseract is an open source OCR. See <http://code.google.com/p/tesseract-ocr/>

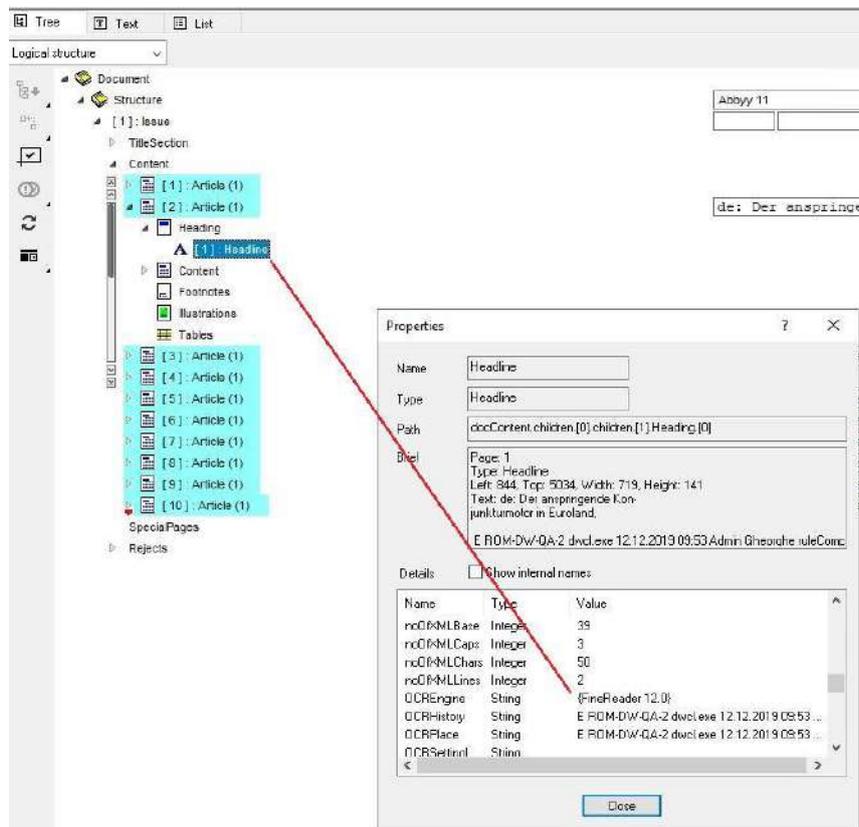
See [docWizz Reference Book](#), [Tesseract OCR](#) for configuration details.

Tesseract can be used as main OCR and the client/services will start with Tesseract. If Finereader is used as main OCR the user can select Tesseract OCR in the interface (Step:Structure, Task: Review structure and text, select a zone, context menu: Actions, Do Antiqua OCR Built-in) to use it for certain zones.

Example: The main document is in English and only some zones are in Chinese. Then you may use Tesseract to read OCR of the Chinese zones. Or if even a few zones are in Fraktur and you do not have the Finereader Fraktur license available you may use Tesseract for this zones.

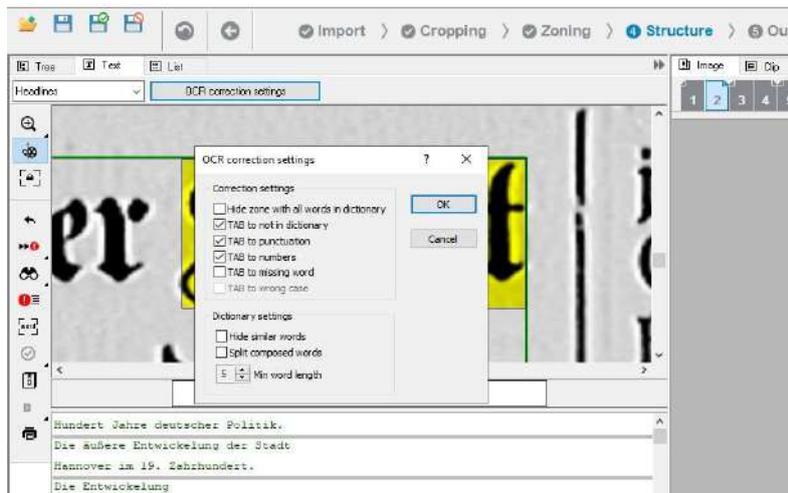


A right click to open the in context menu. Select **Properties** shows, among others, the OCR license used.



OCR correction settings

Note: This button is only available in left hand side working window.



Hide zone with all words in dictionary

Will only show the zones with at least one word not in dictionary (such as hyphens). Activating this may cause the risk that you would not "know" that there are more zones. When clicking on image view on right hand side on a zone which is "hidden", you may forget that and see it as error. So if you want to check or change something again, you need to go to the settings and change it. The text is grey, the characters in question pink, to allow you to check them quickly.

By default the **TAB to punctuation** and **TAB to numbers** are enabled. Both options are working only with **TAB not in dictionary**. Regular text is dark green, any words not in dictionary or below your specified minimum word length are pink. TAB will move cursor to the next word which is not in the dictionary. That will allow to doublecheck only the words unvalidated by dictionary, the other words being skipped.

Example:

"everyone" is in dict, "everyone's" is still showing as green.

Whole words are in dictionary, everyone's is a punctuation case that tab will stop to (but *everyone* is the word in dictionary, *everyone's* still needs to be verified)

Selecting *everyone's* and putting it to dictionary will add again *everyone* (punctuation is trimmed on AddWordToDictionary).

To hide punctuation for words in dictionary uncheck **TAB to Punctuation**. All the punctuations will be considered as not being part of dictionary and the correction browsing will include all of them.

To hide numbers uncheck **TAB to numbers**. All regular words will be greyed out, and using TAB will take you to the next number (in pink). All the numbers will be considered as not being part of dictionary and the correction browsing will include all of them.

TAB to missing word: All text greyed out, except words that were not found in the dictionary. These have the first or last letter highlighted in light blue. Based on the zone width, font size and the number of characters are identified the positions in the text where the OCR could have been failed and skip a word from detection. These positions are indicated with blue and they are included in the correction by checking this box.

TAB to wrong case: Currently not implemented.

Hyphenated words "relationships" => relation-ships" are shown to be added for DICT as "relation" and "ships"

Hyphenated word at the end of the line: relationship is added in dictionary, but for that hyphenated case it will be displayed as in dictionary (it is a draw that when adding you have to add the whole word). Hyphenated words are gray if the whole word is in dictionary.

Hide similar words means that paronym words (similar words core / care) will be displayed as in dictionary or not, this is for the QA user to correct, because even if the words are in dictionary they might still be wrong for that zone (wrongly detected or written). In other words: two words in which one character is different, but all the others identical, are considered similar. When words are added in the dictionary this condition is tested (if is already another word similar in the dictionary) and an attribute for words which have a similar pair is set in the dictionary. By checking this box all the words found in the dictionary as a similar word are considered valid words and will be skipped from the TAB browsing.

Split composed words: TAB will take you from the first half of a compound word to the second half after the hyphen (instead of treating it as a single word). The composed words are split in two at the hyphen (word1-word2 => word1, word2) and each part is validated by the dictionary.

E.g. *Rue-Grand* into *Rue* and *Grand* as two words to be checked in dictionary so that they will be hidden if both of them are in dictionary.

It refers to the possibility to show them as in dictionary or as not in dictionary, For this case, paronyms/words alike.

The **min word length** option can be set from 1 - 99 via readonly scroller. You can adjust, which words will be shown in the text editor in gray color.

Words with at least the given number (e.g. 4 letters) are shown with gray dictionary color. Our recommended length is 5 characters. It set's a benchmark from where the words are starting to be validated by the dictionary. All smaller words are considered invalid.

Due to the **min word setting**, it does not matter anymore what length have the min words inserted in dictionary.

Example:

Set "5" as "min. word length": Every word with at least 5 characters will be shown in gray (e.g. "company"), words with less (e.g. "car") will still be shown in green.

Set "3" as "min. word length": Now also "car" will be shown in gray.

Project based / customized dictionary

Dictionaries can be customized project based for better dictionary management. Operators can fill dictionary and supervisor can double check.

Precondition: Dictionary database and settings in docWizz-global.ini section (DICTIONARY).

Note: The dictionary database has to be configured by CCS.

In the user interface there are default actions to add or remove words from the dictionary in text selection actions from the Text view on the left pane.



- All projects will write into the same dictionary.
- Same entries can be used by multiple projects.
- Select words manually or select all words
- Import existing dictionaries
- Possibility to generate new dictionaries out of corrected zones (e.g. whatever headlines are exported will go to dictionary) either by Extra Tasks or during export.

3.4.5.2.3 Subtask: Review metadata

Metadata can be a book's title, author, subject matter, and a brief plot synopsis along with an abbreviated alpha-numeric identification system which indicated the physical location of the book within the library's shelves. Such data helps classify, aggregate, identify, and locate a particular book.

For example, a digital image may include metadata that describes how large the picture is, the color depth, the image resolution, when the image was created, and other data. A text document's metadata may contain information about how long the document is, who the author is, when the document was written, and a short summary of the document.

docWizz writes all metadata of the document, its issues, chapters, contributions, illustrations and tables to the metadata section. Here, the metadata can be verified, corrected or edited, for example if alternative text for illustrations is of interest.

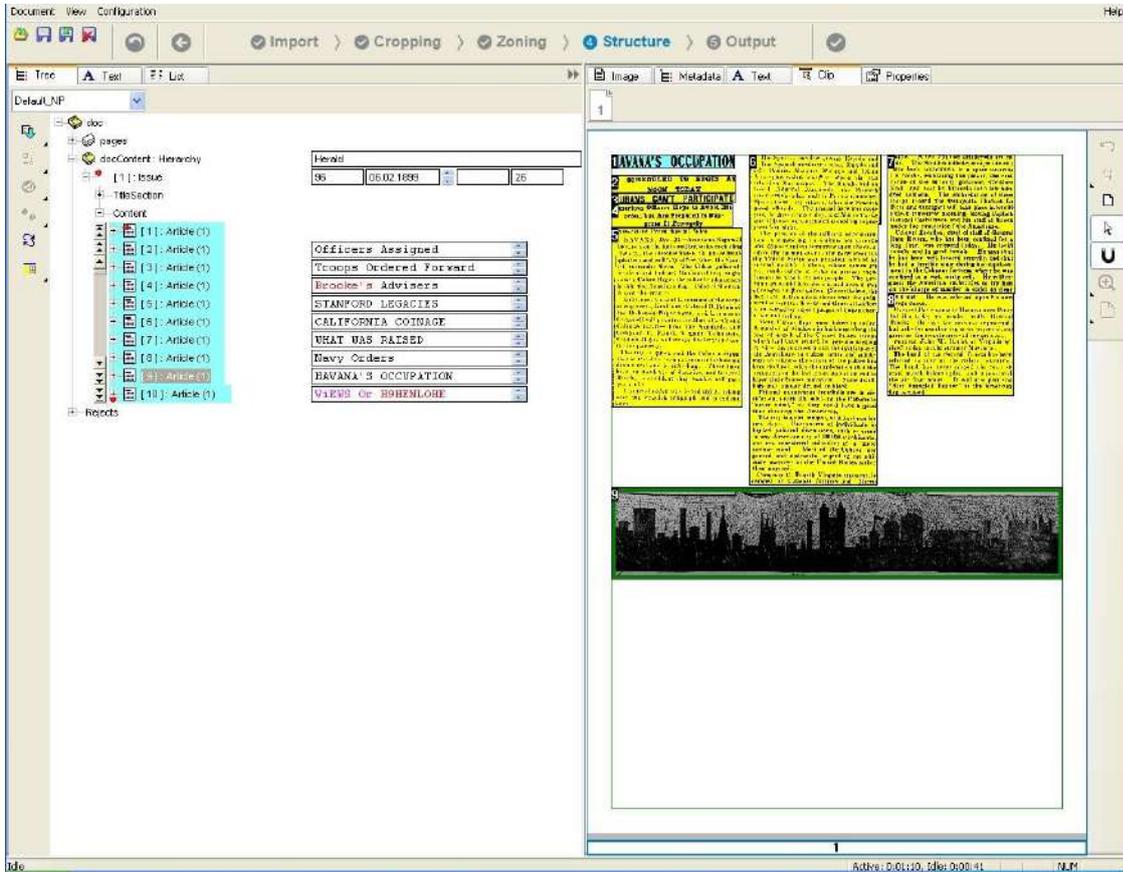


3.4.5.2.4 Subtask: Review clipping

Whole newspaper stocks can be converted and parallel separate documents ("Clippings") can be produced for each article or each other structural component.

After the tasks Recognize layout, Review zoning, Recognize page sequence, Review page sequence or Review structure and text the article structure is detected. It can be defined which structures should be clipped. This is based on TCL script and can be defined for all articles, for all sections, all preambles or any other structure.

Clip single articles by selecting an article in tree (left hand side). You may use the tools on right hand side to sort article zones, cut zones or add pages.

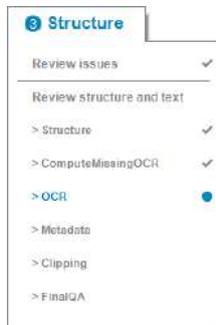


Find the detailed description of the tools: [Clip view](#) .

3.4.5.2.5 Special feature: Sub tasks

This task is deactivated per default.

An additional QA step can be implemented that allows administrators, team leaders/project leaders to do a final QA on documents. In this scenario, we distinguish between normal users and QA users. Normal users do correction on the documents as usual. QA users perform a final check on the whole document before it is exported. In this chapter you will learn how to work with the QA mode.



Depending on which subtasks you have enabled, these are displayed in the workflow drop-down list in the STRUCTURE step

Completed subtasks are marked with a check, the current subtask has a blue dot

Processing a document forward/back between subtasks:

click on the next task or the "Process" button to send the document to the next subtask,
or click on a previous subtask to route back the document

The order of the subtasks cannot be changed, however, not all of them need to be enabled. The choice is yours – select the tasks that you want to make mandatory

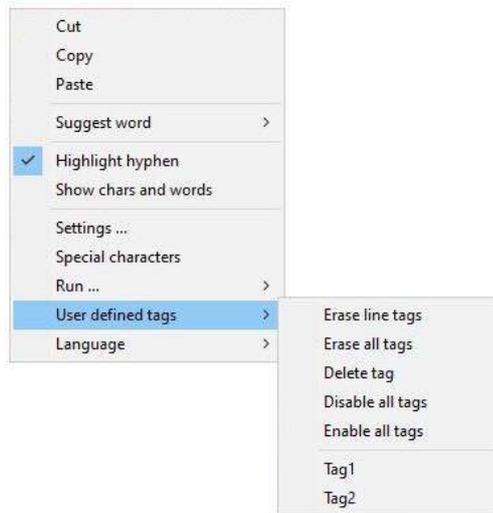
The subtasks are project-based, i.e. they only apply to the project that you have specified, but need to be fixed in the project configuration before your project starts!

See detailed explanation **14-dW_SubTasks_3-3.pptx** on CCS information and support center.

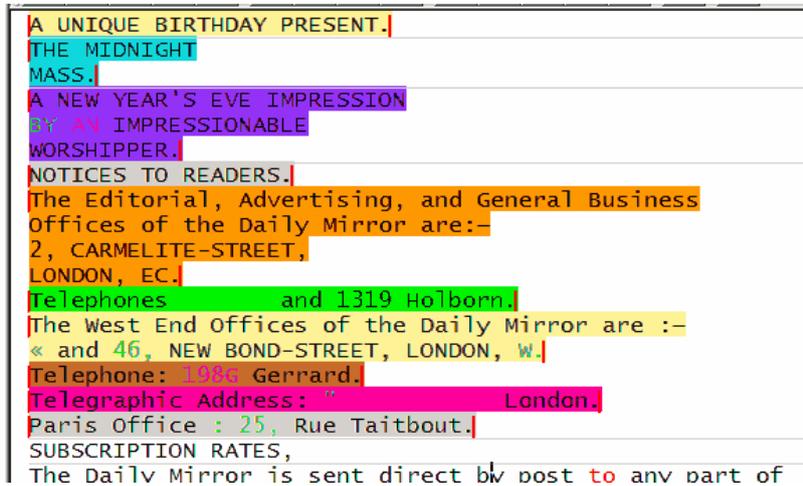
3.4.5.2.6 Special feature: Tag on text

Adds tags into XMLtxt like objects allowing text be logically separated into intervals. Different tags are visible using configured colors.

As you can see the user defined tags are available in the contextual menu, and below set on text.



More than 10 tags can be entered in Book-DW.xml (or newspaper) and are shown. Shortcuts only work on first ten.



By default all tags are visible (are drawn on text) but you can hide any of them: ensure there is no text selection and click in menu (uncheck tag menu). If there is no selection menu shows/hides selected tag.

Be sure you don't have selection otherwise you insert/change selected tag.

Tags are delimited by a red vertical line, this way user can be sure of the boundaries of a tag, having two tags of the same type one near the other, would have been difficult to differentiate.

To resize a tag simply insert that tag over. From GUI you can remove all tags from current line or from all text.

EraseLineTags option from contextual menu, will remove tags that are found on selected line. Note: if tag is spread on more lines, it will be erased.

EraseAllTags option from contextual menu, removes all tags that were assigned on current tree selection (if you use selection on imagePage for example), only the tags on current imagePage selected in tree will be erased, the other tags on other imagePages, will be kept.

DeleteTag option from contextual menu, deletes only selected tag. Note, if you select an overlapped area of two tags, both tags will be deleted.

Inserting, removing and show/hide tags are possible using shortcuts as well: (Ctrl+Shift+1) ... (8)

Shortcuts are recommended way of inserting/removing tags! Menu is recommended way to show/hide tags.

Note: "TagOnText" and "DICT" features are mutually exclusive. By software design they are not meant to work together. This fact is due to the highlighting by colors and while one is related to layout and the other to OCR these two should not be needed at the same time. You can not draw too many overlapped colors one over the others. They will not work both at a time, you either have one, or the other. Tag on text is available everywhere but DICT only on Outline frame. One is related to layout correction the other is related to OCR correction.

Overlapped tags

In this case we have three tags:

- **First Tag:** NOTICES TO READERS The Editorial,
- **Second Tag:** The Editorial, Advertising, and General Business Offices of the Daily Mirror are:- 2, CARMELITE-STREET.
- **Third Tag:** CARMELITE-STREET, LONDON, EC. Telephones : 1310 and 1319 Holburn.

A NEW YEAR'S EVE IMPRESSION
BY AN IMPRESSIONABLE
WORSHIPPER.

NOTICES TO READERS.
The Editorial, Advertising, and General Business
Offices of the Daily Mirror are:-
2, CARMELITE-STREET,
LONDON, EC.
Telephones : 1310 and 1319 Holborn.
The West End Offices of the Daily Mirror are :-
« and 46, NEW BOND-STREET, LONDON, W.

When tags get overlapped the overlapped color is a mixture between the colors of the two overlapped tags.

NOTE: you can have more than two overlapped tags, or a mixture between overlapped and imbricated tags.

Imbricated tags

- **First tag:** From NOTICES to Rue
- **Second tag:** From CARMELITE to 46

Post processing or user action, after tags have been generated, automatic hierarchy can be generated based on the tag type as in the below screenshot: Topic with its all hierarchical contents.

The screenshot displays a hierarchical tree on the left side of the interface. The tree structure is as follows:

- 1997-98
 - I. Budsjett
 - 111. Spørsmål
 - IV. Lovsaker

The right side of the interface shows a text editor with the following text and tags:

1997-98 med en 100 pst. ufor mann, sett i lys av at bl.a. skilte og separerte får slik utvidelse, S.tid. 1236-1237 (17.12.97).

IV. Lovsaker:

I. Lov om endringer i lov av 24. oktober 1946 nr. 2 om barnetrygd. (Automatisk tildeling av barnetrygd; opphevelse av 30-dagersregelen for nyfødte som dør.)

Ot.prp. 47, ref. O.tid. 250 (23.4.98), Innst. O. 52, O.tid. 300 (19.5.98), Besl. O. 57, L.tid. 32 (4.6.98), Lov av 18. juni 1998.

3.4.5.2.8 Implementation notes and limitations

- By default, if NER (Named Entity Recognition) is set for a particular document, the specific tags are available, else there is no other functionality without custom configuration. More details about NER can be found on [NER specific tags](#) subchapter.
- No way to use at the same time dictionary and tags. This protection is added only to outline text as only there are dictionaries.
- Only first 8 configured tags have shortcuts, the rest of them are available via contextual menu (shortcuts (Ctrl+Shift) + number{1-8})
- Copy/cut/paste are limited in functionality.
- Tags depend on the order in which the text is inserted into text editor: Following tag starts on text block number 4 and ends on 5. If select 4 and 5 tag will show but if selecting 5 and 4 the text will be inserted into control in reverse order so although start and end tags has the same name, editor will consider the two tags as being 2 successive tags of the same type not a single tag interval.

The screenshot shows a text editor with the following text:

to prepare the materials of four brigant were to be launched into the South Sea. The timber was felled on the Atlantic seaboard; was then, with the anchors and rigging, transported across the lofty ridge of mountains opposite shores of the Isthmus. Several Spanish guards, thirty Negroes, and a great number of Indians were employed for the purpose. They had no other roads but Indian paths, straggled through almost impervious forests, across the and up rugged defiles, broken by rocks and

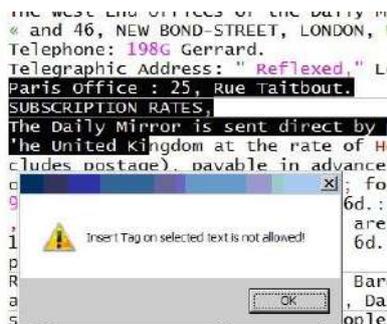
On arriving at Acla, Vasco

The status bar at the bottom indicates: Line: 1, Column: 1, Language: English, Active: 0:01:56, Idle: 0:00:2...

- Editing text into both left and right text views is **strongly NOT recommended!**:

Because text editor computes tags over the same text before displaying into each control, drawing fails.

- Tag on more than one zone is not allowed, tag should be on a single zone selection:



- Tags can't be made on a part of a word. They are designed to work on entire words. If you try to make the tag on a partial word, it will not be made, if you make a longer selection but ending on a partial word selection, the tag will be marked until the beginning of the partial selected word. See selection was made until *cross*, which is a partial word selection. The tag is marked until the start of that partial word.

,~ subscribers abroad the terms are :
 16s. 3d. ; for six months, 325. 6d. ;
 payable in advance.
 Remittances should be **crossed** " Barcl
 and made payable to the Manager, Dail
 suggest such a thing; but if people i
 theatres would sometimes imagine to t
 selves what they would do in case of

3.4.5.2.9 NER specific tags

By default, if NER (Named Entity Recognition) is enabled, following tags are set on each text area:

- PER: People, including fictional.
- FAC: Buildings, airports, highways, bridges, etc.
- ORG: Companies, agencies, institutions, etc.
- GPE: Countries, cities, states.
- LOC: Non-GPE locations, mountain ranges, bodies of water; for some models, GPE is not detected as separate category, and GPE/LOC will be tagged as LOC
- EVN: Named hurricanes, battles, wars, sports events, etc.
- WOA: Titles of books, songs, etc.

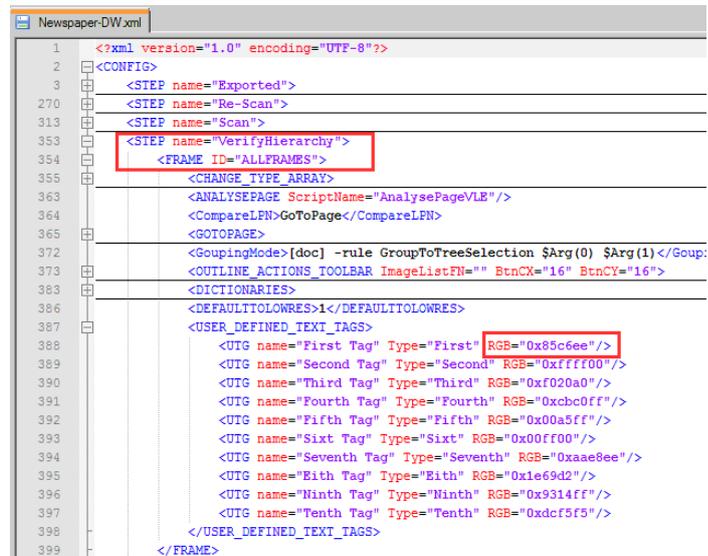
Default color codes and shortcuts used to edit the default tags are:

PER: People, including fictional	- CTRL-SHIFT-1
FAC: Buildings, airports, highways, bridges, etc.	- CTRL-SHIFT-2
ORG: Companies, agencies, institutions, etc.	- CTRL-SHIFT-3
GPE: Countries, cities, states	- CTRL-SHIFT-4
LOC: Non-GPE locations, mountain ranges, bodies of water;	- CTRL-SHIFT-5
EVN: Named hurricanes, battles, wars, sports events, etc.	- CTRL-SHIFT-6
WOA: Titles of books, songs, etc.	- CTRL-SHIFT-7

3.4.5.2.10 Configuration of tags

Tags configuration is stored into `..\config\PVSCFG*doc_type*-DW.xml` so tags may be configured very flexible: on project, doc type, step, both views (outline and detail though not recommended to use both at the same time) if you use **ALLFRAMES** place or finally each frame.

Sample cfg:



```
1 <?xml version="1.0" encoding="UTF-8"?>
2 <CONFIG>
3   <STEP name="Exported">
270   <STEP name="Re-Scan">
313   <STEP name="Scan">
353   <STEP name="VerifyHierarchy">
354     <FRAME ID="ALLFRAMES">
355       <CHANGE_TYPE_ARRAY>
363         <ANALYSEPAGE ScriptName="AnalysePageVLE"/>
364         <CompareLPN>GoToPage</CompareLPN>
365         <GOTOPAGE>
372         <GroupingMode>[doc] -rule GroupToTreeSelection $Arg(0) $Arg(1)</Goup:
373         <OUTLINE_ACTIONS_TOOLBAR ImageListFN="" BtnCX="16" BtnCY="16">
383         <DICTIONARIES>
386         <DEFAULTTLOWRES>1</DEFAULTTLOWRES>
387         <USER_DEFINED_TEXT_TAGS>
388           <UTG name="First Tag" Type="First" RGB="0x85c6ee"/>
389           <UTG name="Second Tag" Type="Second" RGB="0xffff00"/>
390           <UTG name="Third Tag" Type="Third" RGB="0xf020a0"/>
391           <UTG name="Fourth Tag" Type="Fourth" RGB="0xcbc0ff"/>
392           <UTG name="Fifth Tag" Type="Fifth" RGB="0x00a5ff"/>
393           <UTG name="Sixt Tag" Type="Sixt" RGB="0x00ff00"/>
394           <UTG name="Seventh Tag" Type="Seventh" RGB="0xaae8ee"/>
395           <UTG name="Eith Tag" Type="Eith" RGB="0x1e69d2"/>
396           <UTG name="Ninth Tag" Type="Ninth" RGB="0x9314ff"/>
397           <UTG name="Tenth Tag" Type="Tenth" RGB="0xdcf5f5"/>
398         </USER_DEFINED_TEXT_TAGS>
399       </FRAME>
```

- File to use for tags configuration
- Here you configure all text views for Review structure and text
- Tag color. **Although color is named RGB actually the colors are read as BGR!** So for example "First Tag" has following color channels: R=0x85, G=0xC6 and B=0xEE. If this will create confusion in future, this will be renamed either to BGR or the implementation will be changed to accept exactly RGB.
- "name" attribute is used internally in code **only "Type" attribute is used in GUI**. Both may have the same value.

3.4.5.2.11 Special feature: PDF vote

PDF vote - Compare original text from a PDF with OCR result. OCR Accuracy will be improved through "Voting" technology.

dW processes different types of source documents. Beside scanned paper document with the scanner documents are processed, which are already present in an electronic form. If so called *bitmaps* are processed, thus pictures build up only from pixels, the same is present as after scanning (Tiff, JPG etc.).

With the import of external documents, which not only contain the picture but also the text of the shown document (PDF, DOC), it is not naturally necessarily to take the text from the picture and run OCR again. Apart from the expenditure of time, the determination of the text via OCR necessarily also inaccuracies of the OCR are to be expected. Errors in the determined text may appear, too.

With the concept of the PDF import a function was already implemented to take over the text in such a way that no more OCR processing is required.

The variability of the PDF documents which can be imported led to the realization that the text picked out from the PDF is not always the text, which is represented by Acrobat reader on the document page.

Different causes lead to the fact that the text is not useful or missing:

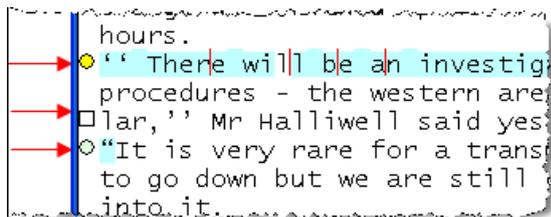
- The text is deposited also in the PDF document only as bitmap (thus as picture), is not present at all in text form.
- With the production of the PDF document corrections were made, which are visible on the screen in the Acrobat Reader, but not in the deposited text. It can happen that complete articles supply another text, as that, which is visible.
- The deposited text is formatted very strong and supplies many blanks within the words or blanks are missing.

The concept of PDF Vote became developed to process articles or paragraphs, with which the deposited text is correct.

In order to get a more accurate text from documents imported from PDF, the PDF's embedded text and the OCR resulted with ABBYY are compared so the best version of the text can be used.

Small **circles** view the **OCR** text.

Small **squares** view the **PDF** text.



For the option to compare the text, the document needs to be imported with a special RDY file: [Newspaper-ocrpdf-comp.xml](#). This RDY file sets the profile of the document as PDF Import, and for OCR uses "OCR-PDFComp.tcl" script.

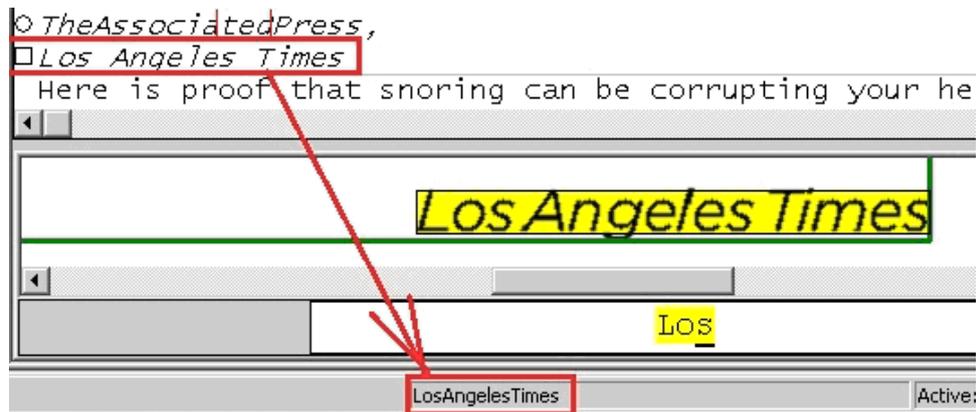
The two versions of text are displayed overlapped (on top – the embedded text), and where differences are found, the different characters are highlighted with a light blue color. If the difference is a missing character on the embedded text, this difference is highlighted with a red line.

Also, in front of the line of text where a difference is found, a pointer is displayed (o). To display the text detected by the OCR ABBYY, you have to press the pointer, and the other text is displayed. Also the pointer changes its shape to a square (□).



In case there aren't any differences between the two versions of text, only one is displayed, with no pointer in front of the text line.

For text lines with differences, when the mouse is hovered over the respective line, on the status bar is displayed the other version of the text.

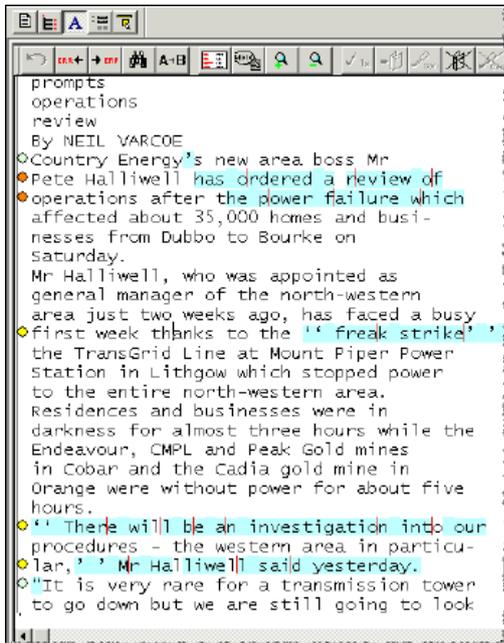


Color codes

Colored dots at the left hand side beside the text indicate the recognized differences. Here the colors mean:

- **red** mark lines in which a difference is recognized and the OCR does not match the PDF text
- **blue** shows differences with blanks
- **yellow** shows differences with punctuation
- **light green** indicates that in this line there are words already corrected by the system
- **orange** shows differences word by word, the words are highlighted
- **white** represents the OCR text
- **white** white square represents the PDF text

Clicking on a color point shows the differences in the current line.



To work faster you can also use keyboard shortcuts:

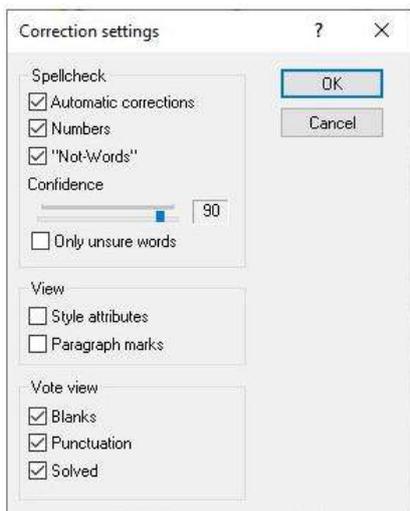
↓ [Alt]	Jump to next line with a difference
↑ [Alt]	Jump to previous line with a difference

If the text cursor is located into a line with such a marking, a change can be achieved with the following combinations of keys:

→ [Alt]	Show next alternative (switch between OCR and PDF text)
← [Alt]	Show the highlighting again

Text correction settings for PDF vote

You can configure text correction to suit your requirements. Open the **Correction Settings** dialog box by placing the mouse pointer on the correction screen and clicking the right mouse button.



You use the **Check** area to specify the types of errors that should be submitted to the user for verification. When you select the options provided under **Check**, the system jumps to the corresponding points in the text in order.

If you select **Automatic Corrections** the system jumps to the words that it automatically corrected and which appear in blue.

If you select **Numbers**, the system jumps to numbers, which appear in green.

If you select **Not-words**, the system jumps to the abbreviations, coded in green.

Using the **Confidence** slider  you can choose on the scale a value between 1 and 90 to select the confidence level that the system should use to correct texts.

If you select **Only unsure words** the system jumps only to words that are unsure.

In the **View** range you can specify the typographical attributes to be displayed in the text window.

Selecting **Style Attributes** means the system will show the text with the original document attributes such as bold text, italics, etc.

Selecting **Paragraph Marks** means the system shows the text as it is organized in the original document.

Within the **Vote View** range you specify whether the marks in the text for **Blanks**, **Punctuation** and **Solved** differences should be shown or not. For this view switching on of the function **PDF Vote** is presupposed, this is adjusted in the *.rdy file.

3.4.5.2.12 *.rdy file entry for PDF Vote

OCRCfg.xml file entry for PDF Vote

The **OCRCfg.xml** file has to contain

```
<SET name="OCRScripts" enable="1">
<SET name="PATH" enable="1">***DATA***\script\OCR-PDFComp.tcl </SET>
</SET>
```

to enable the PDF Voting mechanism.

3.4.5.2.13 Example text correction with PDF Vote

Click into a line, which was marked with a colored dot. Use the shortcut (Alt)  to show next alternative. The word changes between OCR text and PDF text back and forth.

Example 1: orange ●

Differences are shown word by word.

NEWCASTLE City Council
recorded another healthy cash
surplus in the December quarter,
● but money collected from parking
● meters came in at \$500,000 below
what was expected.
○ Despite some of the council's

NEWCASTLE City Council
recorded another healthy cash
surplus in the December quarter,
○ but money collected fl'om parking
● meters came in at \$500,000 below
what was expected.
○ Despite some of the council's

Example 2: blue ●

Shows differences with blanks

east had an impact, she said.
● A greater number of meters were
inoperable late last year through
faults or vandalism, she said.
○ Despite the drop in anticipated

east had an impact, she said.
○ A greater number of meters were
inoperable late last year through
faults or vandalism, she said.
○ Despite the drop in anticipated

east had an impact, she said.
○ A greater number of meters were
inoperable late last year through
faults or vandalism, she said.
○ Despite the drop in anticipated

Example 3: yellow ●

Shows differences in punctuation

storms dropped heavy rain on a few
● inland areas including Trangie (34mm),
● Coonamble (16.4mm) and Lightning
● Ridge (34mm).
● Currently, 93.2 per cent of NSW is
drought declared.

● inland areas including Trangie (34mm),
● Coonamble (16.4mm) and Lightning
○ Ridge (34mm).
○ Currently 93.2 per cent of NSW is
drought declared.

Example 4:

(Alt) → The whole block is taken from the OCR result.

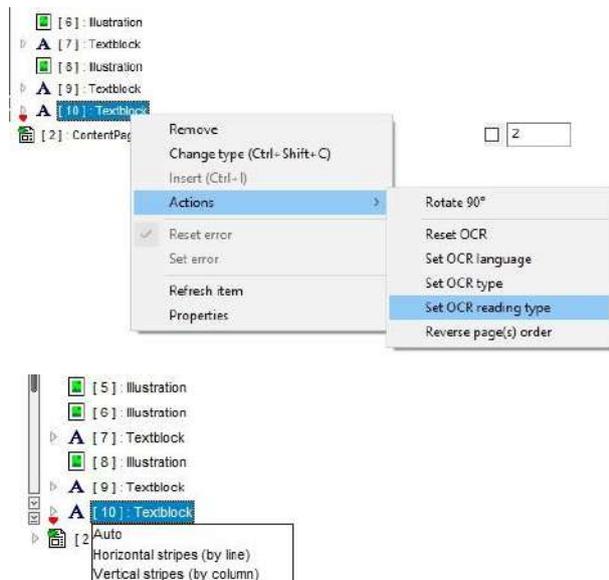
3.4.5.2.14 Special feature: Languages with different reading rules

Some languages have different reading rules, meaning that text is read line by line, or column by column (vertical text).



There are languages where one single option is available (f.e. european languages - horizontal, left-to-write, Hebrew - horizontal right-to-left) but as well languages where both vertical/horizontal reading is used f.e. Japanese/Chinese.

Some languages have different reading rules, meaning that text is read line by line, or column by column (vertical text).



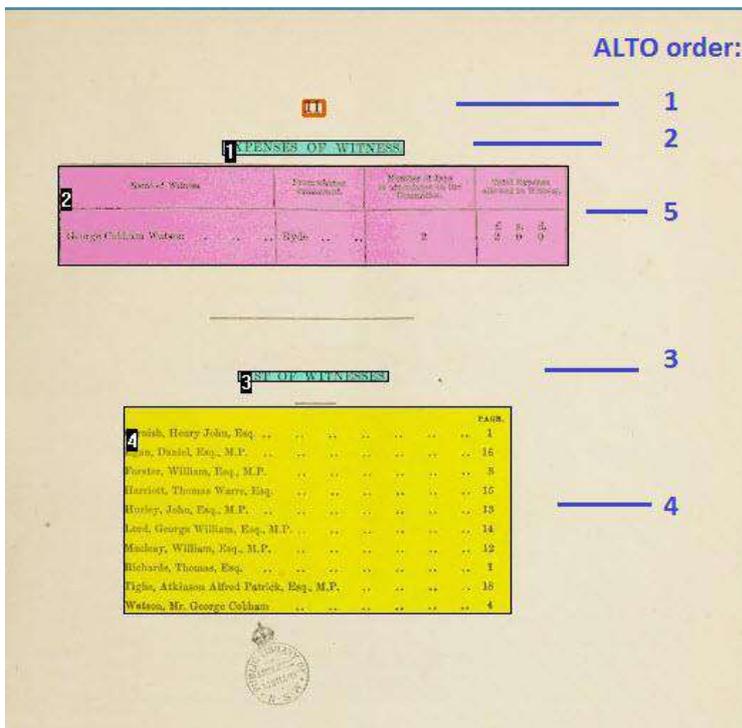
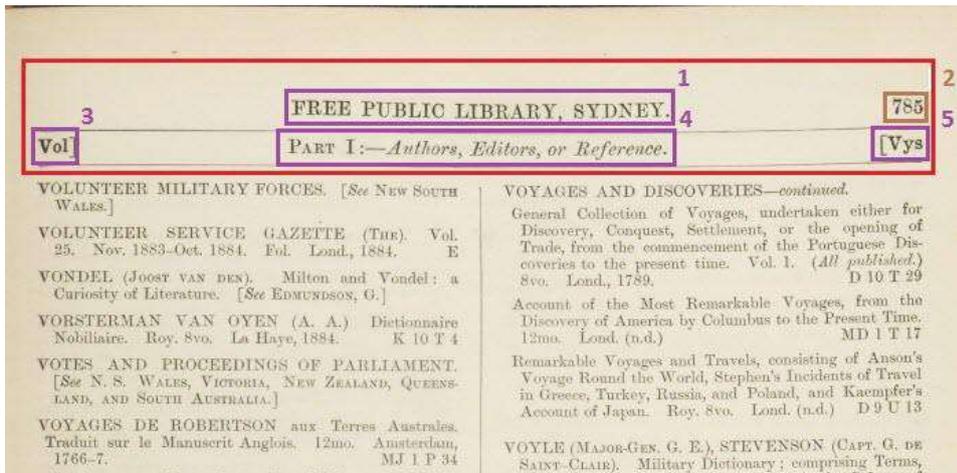
Abby FR engine tries to detect automatically which is the case. Now is possible to change in interface, for a zone, or more zones the way the text is read: auto detection, horizontal stripes (by line), vertical stripes (by column).

It is possible to force this via project configuration (OCRCfg.xml).

3.4.5.2.15 Order of zones in ALTO

docWizz use the following rules to sort zone in an ALTO file:

- If Structure is defined (for project where structure is required) the order of zones is given by "reading order" according with the structure. Please note that Tables/Illustrations/Advertisements are usually put at the end of a Chapter, so they will most likely will appear at the end of the page in ALTO. If these need to be into the exact position, please use specific types "GraphicalIllustration"/"GraphicalTable" which can stay in Content, beside Paragraphs. In this way you force the reading order to take these into account.
- Zones not in structure (like page numbers, running titles) will stay usually on ALTO margins (TopMargin/BottomMargin) and will be sorted into those margins based on left/top rule. This do not guaranty 100% precise results since "order" is not mathematically defined in a bi-dimensional space, but will provide meaningful results in majority of the cases.
- Page Level projects - order is defined based on rule "left/top" reading order. This do not guaranty 100% precise results since "order" is not mathematically defined in a bi-dimensional space, but will provide meaningful results in majority of the cases.
- Running titles/page numbers are always on top/bottom margin and sorted top-left (grid based).



3.4.6 Output

Final checks are done in the Output step.

3.4.6.1 How to use the Review output task

Once the document has reached this step, METS and ALTO files, and any other configured output types are already created and saved in the Output directory docWizz/Out.

In the Review output task you check the final output. No change is allowed in this step. If any problem is encountered, the document must be returned to a previous step to be corrected.

After you have performed all necessary and desired checks and corrections, finally press the process

button  a last time.

docWizz creates a backup of the document and places it in the Backup directory `dwShare/BACKUP`, and the document is deleted from the Document pool `docWizz/Pool`, and the In `docWizz/In` files remain.

After successful processing the current document will be deleted from the docWizz user interface.

3.4.7 Rejects

A reject within docWizz indicates that a document is not accepted or not valid. A reject can be manually set to a document by the QA team. When the document is rejected, it is returned to the operator. A reject can be an automated reject for which the system has detected an error in the document. An operator must either correct the reject or accept it. There exists different rejects for each step. Rejects are configurable in the project configuration.

See [docWizz ReferenceBook](#) for configuration details. Chapter [Automated QA: Reject Conditions](#).

Rejects manager

Rejects have their own dialog, which can be kept on the screen without having to switch between List, Tree or Text view.

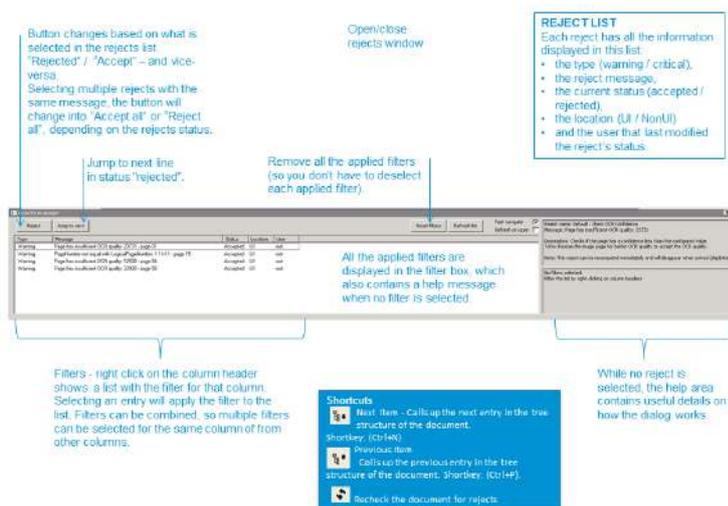


Figure 3: Rejects manager

Reject list

The Rejects dialog has a list on the left side that contains all the rejects; each reject has all the information displayed in this list: the type (warning/critical), the reject message, the current status (accepted / rejected), the location (UI / NonUI) and the user that last modified the reject's status.

Filters

Each column has filters: when using right click on the column header, a list with the filter for that column is displayed. Selecting an entry will apply the filter to the list. Filters can be combined, so multiple filters can be selected for the same column or from other columns.

Filter-box

All the applied filters are displayed in the filter box, which also contains a help message when no filter is selected. On the right hand side of the filter box we have the Reset button, which removes all the applied filters (so you don't have to deselect each applied filter).

Help area

Below the filter box we have the help area. While no reject is selected, the help area contains useful details on how the dialog works. As soon a reject is selected, the content that was previously displayed in a tooltip is now displayed here.

Accept button

On the left hand side of the dialog we have the acceptance button – this button changes based on what is selected in the rejects list. If you select an element with "Rejected" status, the button will be "Accept" – and vice-versa. Selecting multiple rejects with the same message, the button will change into "Accept all" or "Reject all", depending on the rejects status. Of course, this only applies to warnings, not to Critical rejects.

Jump to next button

Jump to next line in status "rejected".

Additional functions

We have some additional functions in the middle part of the dialog:

- "Refresh list" – of course, re-computes the rejects list
- The "refresh on open" check – this can be used if you want the rejects list to be re-computed each time the dialog is opened
- "Fast-navigate" check – this check will make the dialog a little bit more dynamic – for example, accepting a reject, the next one is selected automatically. Also, the text area or zone is automatically selected, and the view will be changed in the event that that the text area or zone is only appearing on a different view, such as the "Complete Document" view.

Docking

A special feature for this dialog is that it can be moved anywhere on the screen or on a second screen, and also, it can snap back into place (on the bottom part or on the top part of docWizz). The last position of the dialog is kept.

How to open

The dialog is automatically opened when processing a document and rejects are present, when selecting "Rejects" entry from List view or when pressing the "Rejects manager dialog" button, available in Tree and List view for tasks with image view on the right side, and on Image view toolbar for the other views.

Disabled List view

The List view was disabled for Rescan, Setup import, Review import, Prepare cropping and Review cropping; it was used in this steps only for fixing / verifying rejects. Since the Rejects Manager was implemented, there is no need for this view.

Automatic rejects

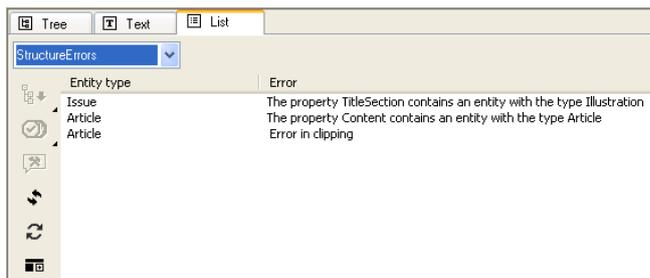
Automatic rejects are calculated when processing from one step to the next step. Rejects can be accepted if they are not critical.

If a reject is critical, then it must be corrected before processing to the next step. Identical reject type can be selected at the same and accepted together.

Reject status contains the reason why the reject was raised. User action tells which rejects have been accepted or not.

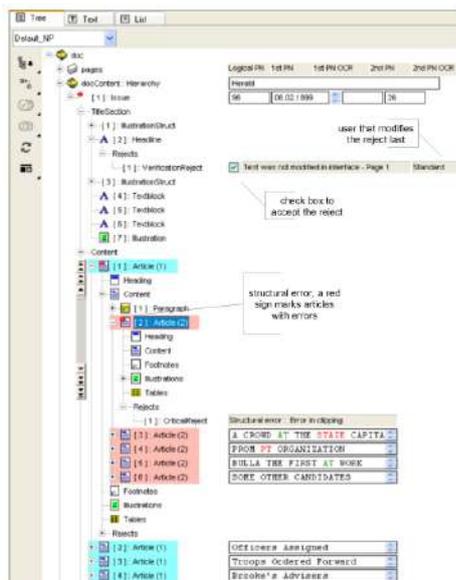
Remark: If this list is empty this can happen because document has not been saved yet in Import step, after save and re-opening the document, the rejects appear in List view.

StructureErrors in List view contain also all the entities that have rejects that are not accepted. The Error column holds the reject reason for that entity.



Some rejects are defined to be accepted from the user interface in the tree view.

Rejects can be seen here as well, along with status (checkbox), user and description.



Beside this you can see the reject reason and also the last user that changed the status of the reject.

The tooltip for rejects displays what the reject checks and how to fix the reject.

Also, for the rejects that are computed only by services ("SkipInteractive = 1"), this message will be displayed: "This is a non-GUI reject, for performance reasons will not be re-evaluated in GUI, but only in services processing".

Tool tip for a reject computed in interface:

The screenshot shows a software interface with a 'Rejects' window. At the top, there are tabs for 'Tree', 'Text', and 'List'. Below the tabs is a dropdown menu labeled 'Rejects'. The main area displays a table with three columns: 'Entity type', 'Description', and 'Status'. A tooltip is visible over one of the rows, providing additional details for that reject.

Entity type	Description	Status
VerificationReject	Text was not modified in interface - Page 1	Accepted
CriticalReject	Structural error : Error in clipping	Rejected
VerificationReject	Text was not modified in interface - Page 1	Accepted
VerificationReject	Text was not modified in interface - Page 1	Rejected
VerificationReject	Reject name: Default : check for uncorrected OCR	
VerificationReject	Message: Text was not modified in interface	
CriticalReject		
VerificationReject	Description: Checks if the OCR text was corrected in interface	
VerificationReject	ToDo: Verify if the OCR was detected correctly and accept the reject	
VerificationReject	Note: This reject can be recomputed immediately and will disappear when solved (skipInteractive= 0).	
VerificationReject	Text was not modified in interface - Page 1	Rejected
VerificationReject	Text was not modified in interface - Page 1	Rejected
VerificationReject	Text was not modified in interface - Page 1	Rejected
VerificationReject	Text was not modified in interface - Page 1	Rejected

4 docWizz Control Center



The dWControlCenter is a cockpit for managing the production workflow and system environment. Here you monitor the docWizz services. Steps and tasks can be prioritized and different administration tools are available.

In contrast to a patchwork of different tools as often used in digitization projects, dWControlCenter integrates all necessary components. It enables multi-project management and statistical analysis and covers support- and error-handling. It's highly beneficial for project managers, team leaders and IT administrators.

Log in to dWControlCenter with the same user login mechanism like in docWizz.

4.1 Configuration tool

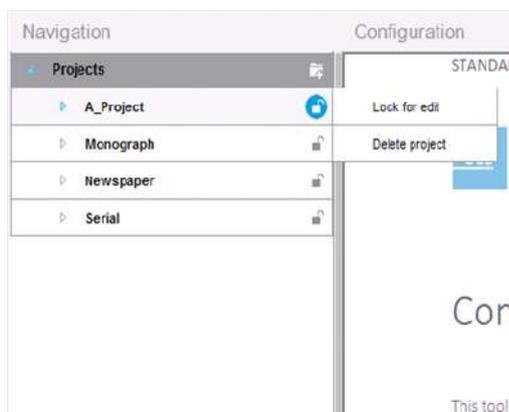
The configuration tool shows an user friendly interface and gives a full overview of your project configurations.

Project configuration files are configured with check boxes. All relevant explanations are actively displayed.

How to create and edit a project configuration

Create a new project in two easy steps:

- Press "Create new project" button
- Add the desired name and document type



Edit the project:

- Hover over the "Lock" icon of the project
- Press "Lock for edit"

Note: The project will be set in “edit mode”; so that no other user will be allowed to make any change to this project until it is unlocked. Also documents that are locked, will not be processed further until unlocked.

After locking a project, new actions become available when hovering over “Lock”:



- Save project
- Discard changes
- Unlock

And two extra buttons:

- Add new settings file
- Delete settings file

Configuration controls

The interface contains easy-to-use controls, making project editing fast and simple:

- Checkboxes
- Dropdown lists
- Edit boxes
- Collapsible groups
- Other...

And a special feature:

- Load from other project – copies the settings from the selected project to the current project

Help at every step

- Help actively displayed for current controls
- Restriction messages for disabled controls
- Validation messages

4.2 Import document

Here you prepare folders for import, trigger import of documents and check the import status.

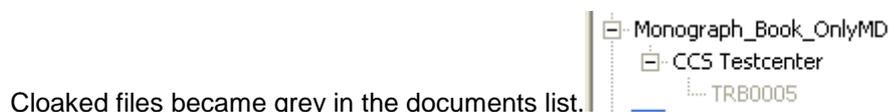
This method to import documents into docWizz is recommended for importing larger amount of documents in batch mode.

This window shows a tree structure. The different projects are shown on top level, what is scanned on second level and the status on third level.

File formats supported: *.tif; *.tiff; *.jpg; *.jpeg; *.jp2; *.pdf; *.cr2; *.png; *.bmp; *.gif to set ready. The same extensions are supported in the import script.

Select a document and click on **Mark for import** button to import the documents into docWizz. If you select a project (top level) all documents on lower levels are also set ready with one click.

There is also a button which is called **Cloak** to create cloaked files to block parsing of current and all subfolders to improve import task performance. Press button to block parsing of current and all sub folders. When a folder has the files "cloaked.rdy" and "cloaked.wrk", the auto-import tasks will not verify this folder and its subfolders for new documents. This can help speed up the task.



Use the **Refresh** button for refreshing the list of new documents that are ready to be imported..

Use the **Set import now!** button to trigger the import task in background. This forces the import tasks and user has not wait for the standard two hours to perform the task. It does not start the import at once.

Use the **Refresh all** button for refreshing without restarting the tool.

It is also possible to store in the **IN** directory a special **Ready** file (*.rdy) in which you can make settings to your needs. docWizz checks this file and processes the files automatically. So you can for example define that the Review import step is skipped.

For more details see [docWizz ReferenceBook](#) manual.

4.3 Services status

Here you manage all the services on root-, group- and services-level.

You can start, stop, kill, shutdown, cancel, and restart services. Another functionality is to check and edit the configuration of services.

If the buttons on the right appear inactive please login as administrator using the Change login button.

When in left tree view a group element is selected, on the right all services are shown as list on the right. By double click (in case you want to apply actions on it) you can jump directly to the service instance on the left tree view.

Service levels

The tree view shows different levels:

The services groups nodes automatically expand in the tree when the numberOfChildren is higher or equal to MaxNoOfEntriesToAutomaticallyOpenTheGroup of [DWControlCenter] section from docWizz-dwsrv.ini

By extending the tree on the left hand side below docWizz Services all machines part of the production environment are shown.

This includes workstations and servers where docWizz service(s) is/are executed.

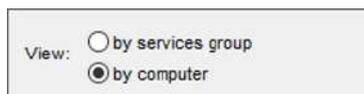
Service icons   will appear on top. The icons were added mostly for the cases when the services are processing really long tasks and the "Stop" / "Restart" commands cannot be completed right away. By adding the icons, it is now easy to see that Stop / Restart action was performed on that service.

-  Stopping will appear when "Stop" command is used on a service
-  Restarting will appear when the "Restart" command is used on a service

docWizz services can be categorized into the following:

- regular services such as: dWSrv, dWSrv2, dWSrv3, and dWSrv4
- special services such as: dWRemoteOCR, dWFTPClient, dWRemoteQAManager, dWRemoteQALoader

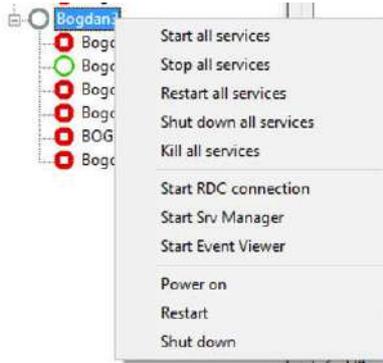
Services can be displayed in tree either by group or by computer.



When services are displayed by computer, additional actions are available on the computer node:

- Start RDC connection
- Start Srv Manager
- Start Event Viewer
- Power on
- Restart

- Shut down



For each machine an icon indicates its current status which can be one of the following:

	Description
	service is executed and currently processing
	service is executed but currently idle. It will pick up documents as soon as there are documents to be processed
	service is stopped and needs to be started in order to pick up documents for processing
	service is about to start
	service is about to stop
	service is about to shut down
	service is about to restart
	service is not available. It either does not exist or has been disconnected. Check local event log/computer is running, but "dWSrvManager" is not running.
	only available on RQA manager. Service is executed on a configuration other than the one of the system (Error Message). Any issues which should be investigated, temporary issues. - In PoolStatus folder (under clients folder) one of the *.csv files (csv.mtn / xml.mtn) has an error into it. On non document error the type is shown as button text: - Error documents on RQA transfer - Missing documents on RQA transfer - Files are too old
	service is performing tasks (e.g. auto import, extra Tasks, ...)
	service is still running, but has not reported back any progress. Is indication that the service might not be operational.
	environment stopped for "maintenance" - service is kept stopped
	environment stopped for "maintenance" - service kept running
	machine is set to night mode

	OCR license expired (RemoteOCR service, Gothic OCR)
	only on Group element - the services within the group have different status. When all have same status, group icon is the same as the state of the services.

By clicking on the according machine one gets a **detailed view** on the selected computer, including current document ID, current job, action performed etc. Here, the according machine can be started, stopped or shut down. Return to docWizz Control Center by hitting docWizz Services on top of the tree again.

Note: On one computer up to 4 instances dW can work parallel. Thus we reach a very efficient use of the hardware and the support of multi-processor computer and Dual- and/or Quad-Core processors.

The Order of the groups at the left tree view is alphabetically.

There are four subtasks (CollectData, CommandFTP, UpdatePool, UpdateReady), how is the handling, when more than one reports a state, which causes the icon to change?

CollectData and UpdatePool both have a different warning/error, which would change the icon on RQAManager in DWCC.

It is shown just the first error is coming up.

Usually all subtasks show the same status like: start, stop, process, maintenance.

In the case of **start** first worker updates icon.

In the case of **stopping** the service the logic is reverse (last child update icon as stop).

In the case of **processing** if any subprocess stats working icon will show RQAManager as working. Subtasks are independent so if anyone is working that icon shows as working.

Monitor progress

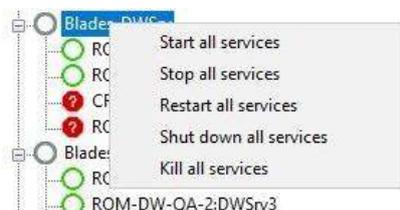
Processing logs (like ID, Task name, Filter, Start time, ...) can be sorted by clicking on header column.

For investigation of systematic issues it is now possible to verify just by one click, if each document in that job failed on that machine. Also possible to filter out f.e. the tasks like AutoDelivery to find just the history of this task and verify the scheduled execution times are working fine.

With right mouse button you open context menus in tree view to start/stop/shutdown/restart/kill... services.

Check now action for services. Because services are not constantly checking if there are new documents or tasks to process, but only once every 2-3 minutes, this action can be used to force the service to search for something to do.

Context menu



Context menu per service

- **Start** - starts the service
- **Stop** - stops the service; task in progress will be finished
- **Cancel** - ends the current process - like pressing "break" when processing a document in step-by-step and most importantly - the service is not stopped. The document is returned to Prepare cropping and is not locked.
- **Shutdown** - task in progress will be interrupted and stops the service. Graceful. Ends (as in waits for a proper moment to interrupt the task, in order to not affect the document) the current task and stops the service. The document is returned to Prepare cropping and is not locked.
- **Kill** - interrupts the task (no matter what) and stops the service. Ungraceful. It will take some time until the dWCC realizes that the service has been killed. So it might take a couple of seconds until the task is displayed in stopped-state. The document is locked in Modify Pages.
- **Restart** - combines a „stop“ and „start“ command

dWSrv is frequently updating the status file (is executed) if the time difference between current time and file time is bigger than 2 minutes. If so, Control Center shows a warning for services having that time difference.

Note: You must not test the time difference when status file is not updated, because then the service could be stopped etc..

There are three colored icons for time difference:



GREEN = No time difference (or less than 2 minutes)



RED = Time difference (greater than 2minutes)



YELLOW = Status undefined (can't be read from remote)

Examples for monitoring FTP and Remote QA processes:

Monitor Remote QCR

Because the Gothic OCR licenses are character limited whenever the remaining number of characters gets very low this needs to be outlined in Control Center properly to make you aware about the status of each Abbyy FR license equipped with Fraktur (Gothic). See **Fraktur (Gothic) OCR licenses**.

There are two warnings implemented:

- one for each service blocking to the machine where this special type of OCR is running
- and a second one on the Remote OCR service belonging to the same machine.

"Hardware information" group on "Services status" tab

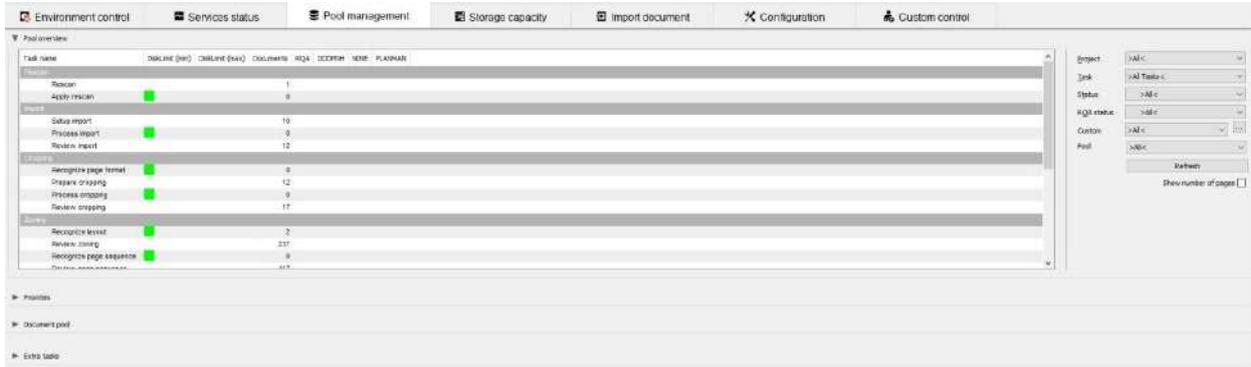
Shown when the selection is on computer node (view: by computer)

Here are displayed all the information about that computer (state on/off, dW applications running, dWServices running, plus all the info about hardware configuration:

4.4 Pool management

Check progress: number of documents/pages per project or task.

Tool tip is displayed showing which limit cause the icon - also for the green ones to show that the green is not due to second condition on same job (f.e. POOL and EXPORT), than for POOL1 and POOL2.



In the area located in the center of the interface the **number of documents for each job** is shown. By hitting the Refresh button the view is updated.

Task name	DiskLimit (min)	DiskLimit (max)	Documents	RQA	DDDPNH	NINE	PLANMAN
Recognize layout	■		2				
Review zoning			237				
Recognize page sequence	■		0				
Review page sequence			117				
Structure							
Recognize issues	■		0				
Review issues			66				
Recognize structure and text	■		0				
Review structure and text			59				
Output							
Prepare output	■		0				
Process output	■		0				
Review output			39				
Cleanup							
Backup / Remove	■		0				
Cleaned	■		0				

Control Center shows number of pages in a different color to make visible, if the user has selected "pages" option.

Project. job, status:

Project: ▼

Task: ▼

Status: ▼

RQA status: ▼

Custom: ▼ ...

Pool: ▼

Show number of pages

You can apply various **filters** in order to reduce complexity of the view. Certain projects, jobs and/or status can be selected from the drop down menus.

The **Project**, **Job** and **Status** combo boxes will behave the same like in Pool open documents dialog. In the list control below, you will see all jobs and the number of documents per job that are matching the selection from combo boxes. If a single job is selected, the list box will show projects and the number of documents. If a single job and a single project is selected, the list box will show each available status and the number of documents.

With the check box **Show number of pages** you can switch between number of pages instead of number of documents.

The Controls will be refreshed from time to time to reduce network/SQL traffic. If you click on **Refresh**, all items are updated immediately.

RQA Client for RemoteQA services.

Control Center displays number of documents per RQA location for automatic jobs in light gray. This helps to see, how many documents will appear soon in which RQA location.

Each process in RemoteQA services reports its processID in the status file. The processID is displayed after the status, like "Idle (PID:5677)" for each RemoteQA service.

Priorities

The list control shows all defined priorities. docWizz will process the documents as specified by the priority settings.



Name	Value	Documents
RESTORE	100	0
RESTOREPOOL	100	0

Buttons: Add, Edit, Remove, Move up, Move down, Refresh list

The list has the following columns:

- Name (JobName/Project/DocumentID/Title/SQL Query)
- Priority **Value**
- No. of **Documents** matching the priority

Priorities

- Priorities are handled from top to bottom.
- The priority value specifies, how often documents of a lower priority are processed.
- A value of 100 defines that first all documents that are matching the priority condition will be processed.
- A value of 80 identifies, that 20% of document processing is used for documents that have a lower priority.
- A value of 0 identifies, that those documents are processed if no other documents are available for processing.
- The single services can be moved up or down or removed from priority list. Existing ones can be edited.
- Priorities will be stored in the document pool database.
- **Move up/Move down** tells the sequence.

Add new Priority

When clicking on **Add**, a dialog opens:

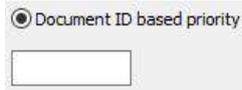


The user may add a **Priority Value** in range of 0-100 (default is 60).

There are different ways to define the priority:

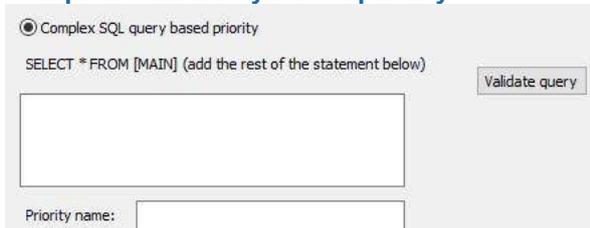
- Task/Project based priority checked – the user selects a job/project that shall run with priority (based on job/project priority) doc priority edit box and SQL query edit box is disabled.
- Doc ID based priority checked - instead of using docWizz, we are also be able to add a doc with high priority from this console. The SQL query edit box and job/project lists are disabled.
- Complex SQL query based priority - and SQL query shall be input. A Validate button is available to be sure a correct query is added. Doc ID edit and job/project lists are disabled.
- Work task

Document ID based priority



Beside the Jobname/Project priority, as well a specific document (ID) can be added. Therefore in pool open dialog in docWizz, the user can perform a right-click on a document and check from context menu "High Priority". Then the document is added at top to the priority queue. Also using Set Priority dialog box this can be done.

Complex SQL Query based priority



Complex SQL query based priority - and SQL query shall be input. A Validate button is available to be sure a correct query is added. Doc ID edit and job/project lists are disabled.

Work task



Handling of locking priorities

"Couldn't lock priorities.xml" appears just in case of a real problem.

Document Pool

The document pool shows intermediate results of documents in any step.

ID	Status	Task	Date	Document Type	Pages	Priority	Group	Title	Label	Project	Workflow	Stop/Job	Shared/Indexed
893	Work	Review cropping	9/30/2020 2:54:00 PM	Register_Monograph-Monograph	51			CroppingRegister...	Register_Mo...	Image Cut an...	Review crop...	0	0
894	Work	Review cropping	9/30/2020 2:54:00 PM	Register_Monograph-Monograph	41			CroppingRegister...	Register_Mo...	Image Cut an...	Review crop...	0	0
895	Work	Review cropping	9/30/2020 2:55:00 PM	Register_Monograph-Monograph	51			CroppingRegister...	Register_Mo...	Image Cut an...	Review crop...	0	0
896	Work	Review cropping	9/30/2020 2:55:00 PM	Register_Monograph-Monograph	41			CroppingRegister...	Register_Mo...	Image Cut an...	Review crop...	0	0
897	Work	Review page sequence	10/7/2020 10:30:00 ...	Register_Monograph-Monograph	139			NOC_008110993...	Register_Mo...	Analysis only	Review zomg...	0	0
898	Work	Review page sequence	10/14/2020 1:51:00 ...	Register_Monograph-Monograph	82	2		BC_Registers00...	Register_Mo...	Image Cut an...	Review page...	0	0
899	Work	Review page sequence	10/14/2020 1:58:00 ...	Register_Monograph-Monograph	59	3		BC_Registers00...	Register_Mo...	Image Cut an...	Review page...	0	0
900	Work	Review page sequence	10/7/2020 10:34:00 ...	Register_Monograph-Monograph	129			NOC_Registers00...	Register_Mo...	Analysis only	Review zomg...	0	0
901	Work	Review page sequence	10/14/2020 1:52:00 ...	Register_Monograph-Monograph	67	2		BC_Registers00...	Register_Mo...	Image Cut an...	Review page...	0	0
902	Work	Review page sequence	10/14/2020 1:59:00 ...	Register_Monograph-Monograph	83	3		BC_Registers00...	Register_Mo...	Image Cut an...	Review page...	0	0
903	Work	Review page sequence	10/7/2020 10:36:00 ...	Register_Monograph-Monograph	91			NOC_Registers00...	Register_Mo...	Analysis only	Review zomg...	0	0
904	Work	Review page sequence	10/14/2020 1:53:00 ...	Register_Monograph-Monograph	49	2		BC_Registers00...	Register_Mo...	Image Cut an...	Review page...	0	0
905	Work	Review page sequence	10/14/2020 2:00:00 ...	Register_Monograph-Monograph	44	3		BC_Registers00...	Register_Mo...	Image Cut an...	Review page...	0	0
906	Work	Review page sequence	10/12/2020 9:17:00 ...	Register_Monograph-Monograph	155			NOC_Registers00...	Register_Mo...	Analysis only	Review zomg...	0	0
907	Work	Review page sequence	10/14/2020 1:54:00 ...	Register_Monograph-Monograph	73	2		BC_Registers00...	Register_Mo...	Image Cut an...	Review page...	0	0
908	Work	Review page sequence	10/14/2020 2:02:00 ...	Register_Monograph-Monograph	66	3		BC_Registers00...	Register_Mo...	Image Cut an...	Review page...	0	0
909	Work	Review page sequence	10/9/2020 2:06:00 PM	Register_Monograph-Monograph	1			BC_RegisterTEST...	Register_Mo...	Image Cut an...	Review page...	0	0
910	Work	Review page sequence	10/9/2020 2:21:00 PM	Register_Monograph-Monograph	38			BC_RegisterTEST...	Register_Mo...	Image Cut an...	Review page...	0	0
911	Work	Review page sequence	10/9/2020 2:23:00 PM	Register_Monograph-Monograph	7			BC_TESTViewTF...	Register_Mo...	Image Cut an...	Review page...	0	0
912	Work	Review page sequence	10/9/2020 3:17:00 PM	Register_Monograph-Monograph	82			BC_RegistersTES...	Register_Mo...	Image Cut an...	Review page...	0	0
913	Work	Review page sequence	10/9/2020 3:36:00 PM	Register_Monograph-Monograph	1			BC_RegisterTEST...	Register_Mo...	Image Cut an...	Review page...	0	0
904	Work	Review page sequence	10/14/2020 1:43:00 ...	Register_Monograph-Monograph	21	1		BC_Registers00...	Register_Mo...	Image Cut an...	Review page...	0	0
905	Work	Review page sequence	10/14/2020 1:55:00 ...	Register_Monograph-Monograph	10	2		BC_Registers00...	Register_Mo...	Image Cut an...	Review page...	0	0
906	Work	Review page sequence	10/14/2020 1:45:00 ...	Register_Monograph-Monograph	24	1		BC_Registers00...	Register_Mo...	Image Cut an...	Review page...	0	0

- In order to give a better overview operators can apply **filters** to show documents in the pool. Please use the drop down menu "Project" to filter by project. Please use the drop down menu "task" to filter by task. Please use the drop down menu "Status" to show only documents that have a certain QA status. One, two, three or none of the filters can be applied.
- Further more, operators can browse for documents within the document pool by typing in the **document ID**.
- The interface disposes of a display showing **number of selected documents** as well as **total number of documents** within the document pool.
- A button **Change Status** has been placed on the right hand side of the pool.
- Sort entries by the arrow on document's list header.
- Operators/Administrators can also enter a reason or comment.

Each document is listed along with its unique **ID**, next **Task**, **Date** of last modification, **Type** (serial or monograph or newspaper) and **Title** of the document. A lock icon indicates a document currently in use.

Whenever a task has been sent to the processing queue, the next step is an automatic process. All these are starting with **Detect...** (exception: SplitDbPages) or **Build...**, so the operator can identify prepared tasks easily.

If an entry starts with **Verify...** or just **Scan** the related document is apparently not prepared for batch processing but waiting for an operator to be opened up.

Status/Labels

Changes Status and/or Labels for selected documents

Status: Work

Label: Work

Priority: QA

Enter your comment:

Comments history:

Comment:

croppingdone

Visible: User Admin

Buttons: OK, Cancel, Add, Copy, Remove, Show all

After selecting one or more documents in the **Document Pool** the status can be changed to a different status by hitting the button.

Note: the status *Correcting on Remote, Remote QA done, Prepared to be sent, Wait for correction and >in use<* are only visible on the remote system.

Reduce functionality is used to free space on pool storage. For reducing storage space in pool, temporary images could be deleted (also cropped images created after MP) and restored if necessary.

- Functionality needs manual actions, just on demand. This is not initialized by workflow dependencies.
- Only administrators can perform the actions while have high impact on pool.
- Tasks must be configured for services. (CCS additional)
- The OnProcess button is disabled if current document is in Restore Pool Data or Reduced Pool Data or Free Pool Data status.
- An image could be restored (e.g. with "Document open") if necessary "on the fly". This will last some time and the user has to wait until document is restored again.
- For safety reasons source images will not be deleted for those pages where the source IN data images are not available at initial path. Only thumbnail images are deleted in this case.
- For restoring the source images must exist in the correct folder (e. g. the IN folder).

In docWizz-dW.ini file RESTOREPOOL, CLEANPOOL tasks must be enabled here:

[PROCESS]

TASKS= ...RESTOREPOOL,CLEANPOOL, ...

The task CLEANPOOL cleans pool data when document status is changed to Free Pool Data, and on completion it sets the status to Reduced Pool Data.

The task RESTOREPOOL restores previously reduced documents to their original pool data.

Set time in for the CLEANPOOL task in <customer>-docWizz-dw.ini file:

[CLEANPOOL.TASK]

PROC=TaskCleanPool

DELAY=300

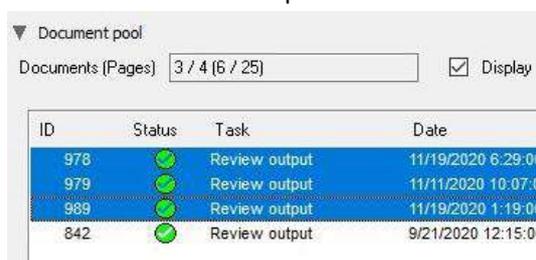
TIME=10:00:00

TIMETYPE=0

...

The manual handling is done in pool dialog and change status dialog:

Select one ore more exported documents.

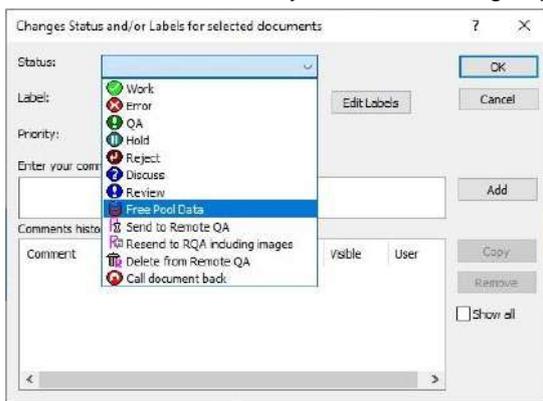


The screenshot shows a dialog box titled "Document pool". At the top, there is a text field "Documents (Pages)" containing "3 / 4 (6 / 25)" and a checked checkbox labeled "Display". Below this is a table with four columns: "ID", "Status", "Task", and "Date". The table contains four rows of data, with the first three rows highlighted in blue.

ID	Status	Task	Date
978	✔	Review output	11/19/2020 6:29:0
979	✔	Review output	11/11/2020 10:07:0
989	✔	Review output	11/19/2020 1:19:0
842	✔	Review output	9/21/2020 12:15:0

Click  button.

Click FreePoolData entry to reduce storage space in pool.



In Control Center you can see then the current active task CleanPool:

In pool view different icons show status of documents:

150371		Exported	01/04/2011 14:29:00	Newspaper	40	86	1130
150374		Exported	01/04/2011 14:29:00	Newspaper	20	86	1130
150377		Exported	01/04/2011 14:30:00	Newspaper	56	86	1130
150378		Exported	01/04/2011 14:30:00	Newspaper	56	86	1130
150379		Exported	01/04/2011 14:28:00	Newspaper	60	86	1130
150380		Exported	01/04/2011 14:30:00	Newspaper	52	86	1130
150381		Exported	01/04/2011 14:28:00	Newspaper	60	86	1130
150382		Exported	01/04/2011 14:30:00	Newspaper	60	86	1130
150383		Exported	01/04/2011 14:28:00	Newspaper	60	86	1130
150384		Exported	01/04/2011 14:28:00	Newspaper	60	86	1130
150346		Exported	01/04/2011 14:28:00	Newspaper	28	86	1130
150366		Exported	01/04/2011 14:28:00	Newspaper	28	86	1132
150369		Exported	01/04/2011 14:28:00	Newspaper	40	86	1132
150372		Exported	01/04/2011 14:29:00	Newspaper	48	86	1132
150375		Exported	01/04/2011 14:29:00	Newspaper	48	86	1132
150385		Exported	01/04/2011 14:28:00	Newspaper	60	86	1132
150394		Exported	01/04/2011 14:28:00	Newspaper	60	86	1132
150396		Exported	01/04/2011 14:28:00	Newspaper	52	86	1132
150397		Exported	01/04/2011 14:28:00	Newspaper	53	86	1132
150398		Exported	01/04/2011 14:28:00	Newspaper	56	86	1132

The  icon shows already reduced pool data. For colored documents about 90% of data can be removed.

Following files will be removed:

- All temporary files b/w images
- Lowres images
- Cropped/aligned images (if they are not changed manually)
- RQA images (always)
- <jobname>.zip (from non-interactive jobs)
- Following files (all non restorable ones) will remain:
- <jobname>.zip (from interactive jobs)
- ID.xml
- rescan images
- deskewed images

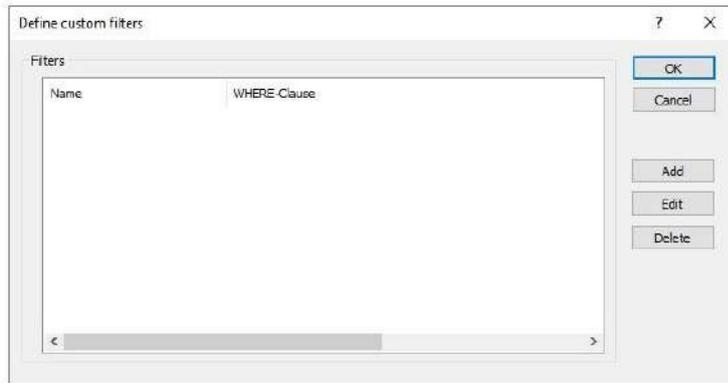
The  icon shows document to be reduced.

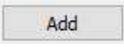
The  icon shows restored pool data. Restores all temporary images data, as were existing before. Uses the task RESTOREPOOL. The document status is set back to the status that was set before the document was sent to reduce data status.

Custom filters:

To have a more specific view on pool, custom filters for Pool dialog are configurable. They contain a pair of displayed name and a fragment of a WHERE expression from SQL select statement. Administrator users can define new filters within UI. Filters will be selectable in a combo box and can be combined with any other filter.

Custom: >All< ... The button with the three dots opens a separate window where you can define custom filters.



Use the  button to create a new filter.



Each document is listed along with its unique **ID**, next **Job**, **Date** of last modification, **Type** (serial or monograph or newspaper) and **Title** of the document. A lock icon indicates a document currently in use.

Whenever a job has been sent to the processing queue, the next job is an automatic process. All these are starting with **Detect...** (exception: SplitDbIPages) or **Build...**, so the operator can identify prepared jobs easily.

If an entry starts with **Verify...** or just **Scan** the related document is apparently not prepared for batch processing but waiting for an operator to be opened up.

The pool folder structure can be extended to two levels of folders to improve performance on mass digitization projects.

When changing a filter manually it is checked whether new document type is available.

4.5 Storage capacity

Here you define disk space limits for different tasks and locations and set "critical disk space" values for "low space" warnings. Services are automatically stopped in case of critical space.

It is not the case that the limit set is the space that we guarantee it will remain free. If the space limit is reached during one document processing then that document processing will be finalized and will fill the limited free space.

Multiplying this case on 10-20 services the space used after the limit is reached can be quite high.

We suggest to set a limit considering the number of services (e.g limit = number of services * 50Gb).

A feature for local export can be customized, if no TEMPFREESPACE node is present in LowDisk.xml, then the default temp space value is considered (2 GB by default).

```
- <MinDiskSpace>
(...)
- <!-- General Limits
-->
<CRITICALFREESPACE Size="15" Unit="K" />
<WARNINGFREESPACE Size="300" Unit="K" />
<TEMPFREESPACE Size="100" Unit="M" />
</MinDiskSpace>
```

It is available in configuration as a comment, and if needed it can be de-commented and adapted by need. This function will not interfere with the disk space check the Control Center does in its UI for export and pool folders.

4.6 Environmental control

Here you check and edit notes people should pay attention to, check and manage error log or create reports and detailed statistics.

Notes

Enter some individual notes. Click first to **Edit** to enable the notes entry area. Click **Save** or **Cancel**.

Error log

The **Error Log** function enables you to refer to the **Error Log** window that automatically lists any errors that have occurred during the current session. In this way, support staff and docWizz administrators have optimal support when looking for the cause of irregularities in the program.

Sometimes, processing fails due to depleted memory. In many of these cases, restarting DWSrv will solve the problem.

If this error occurs, the document will not get error status but remains in the current job to be performed. DWSrv will restart and the document will run again through this job. As soon as the document has 5 or more failures, it will be set to error status anyway.



Restore documents

See a list of documents in the restore queue and restore documents and batches.

Clear log files

Clears the log files.

Set services logon

A dialog box titled "Log-on information" with a question mark and close button in the top right. It contains two text input fields: "User" with a placeholder "Domain\User" and "Password". To the right of the "User" field are "OK" and "Cancel" buttons.

Set user and password.

Volume report

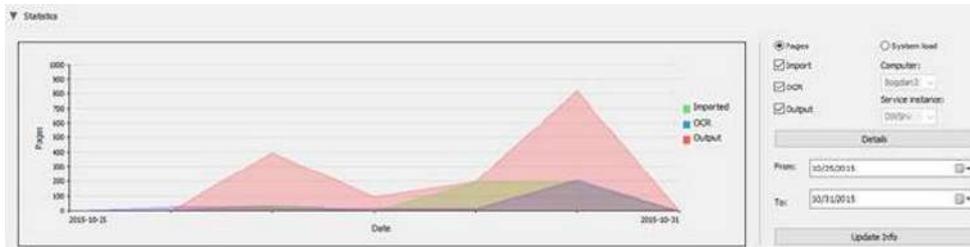
A dialog box titled "Volume report" with a question mark and close button in the top right. It features a dropdown menu showing "11/2020" and a "Create report ..." button. Below this is a text area containing the instruction "Press 'Create report ...' to create report of 11/2020 .". A "Close" button is at the bottom.

By hitting the **Volume Report** button a PDF file containing total page counter and number of pages processed in the selected month is created. For more information please refer to the **Volume report** chapter.

Statistics

With the Statistics tool you get statistical records of the docWizz system. There are different ways to analyze work procedures, jobs or documents. The statistic of docWizz visualizes the logged data about the processed documents and used time, inform about the behavior of the machines and users and so on.

It is used to display graphics regarding the number of pages imported, OCR-ed and output over a period of time. It can also display the System load – the amount of time that docWizz services were processing. This information can be used to take decisions regarding the environment: if the system load is too high, additional services are needed, if the system load is too low, there are services that are not processing.



Backup configuration

Use Backup Config to make a backup of the current configuration. This feature is still under development and will be available in further releases of the Control Center.

What's the VolumeReport?

The VolumeReport is an additional function in docWizz that creates an pdf file containing the total page counter and the number of pages of the selected month.

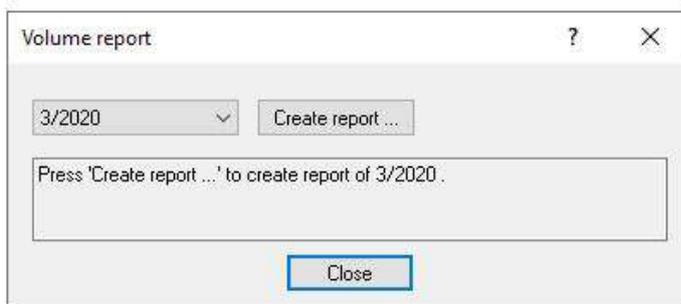
How to create the VolumeReport?

The VolumeReport is positioned at Configuration -> Maintenance. Inside this dialog you can choose "Volume Report...".

A new dialog is opened where you can select the desired period of which you want to create the VolumeReport.

The dropdown box is dynamical filled with the content of `***DATA***\WORK\log.mdb`.

It is selected the minimum date of the BatchResult table and generated all month until last month of current date.



After creating the report by click on the button, the location of the stored PDF-file will be shown in the text area below. The default location of the VolumeReports is `***MAINTENANCE***`. You can change the location in the system configuration in the register "paths". The path name is "MAINTENANCE".

The content of VolumeReport

The VolumeReport contains the following informations:

- short name of the customer
- total page counter
- date of the selected period

- number of processed pages in the selected period
- select completed Pages from BatchResult where date=actualPeriod and JobName='ExportXML'
- two validation codes
- total page counter and number of pages in this period encoded by Base64Encoder
- list of processed pages for each job

Sending the VolumeReport to CCS

At the moment there exists no automatic transfer of these information to CCS in any way. The generated PDF-file has to be send by mail to metae-support@ccs-gmbh.de

How to prove the validation of the VolumeReport

- The properties of the PDF file contains the following information
- title VolumeReport
- creator CCS docWizz
- theme <customerName>
- created at <actual system time>
- application CCS docWizz
- creator CCS Content Conversion Specialists GmbH, Hamburg
- created by docWizz <FVersion> (f.e. docWizz 6.9.0.7)

Behavior in case of errors

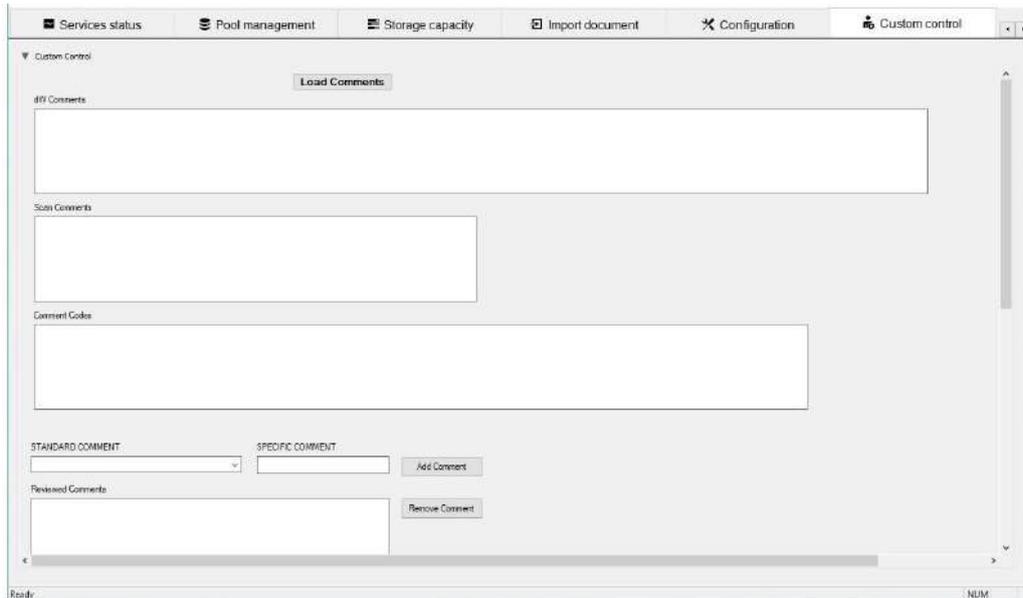
The VolumeReport dialog contains an own filter for errors. If an unexpected error appears the dialog would be closed without a message. Perhaps in the log database an error will be logged. The user can continue his work with docWizz without any problems. In case of expected errors (f.e. error in database requests) the error will be shown and logged in the log database.

Health status group

The environment is monitored and a status (OK, Warning or Error) is displayed for each component:

Name	Status	Message
Services	✘	Error (DWServiceManager is not running on: Bogdan)
Storage	✔	OK
Processing errors	⚠	Warning (494 errors in the last 24 hours)
Document pool	✔	OK (24 documents in error.)
Locked projects	✔	OK (No locked projects.)
RQA status	✘	Error (RQA Manager: Enabled, no processes are running. RQA Loader: Disabled. FTP: 1 stopped.)

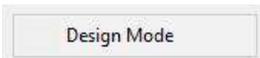
4.7 Custom control



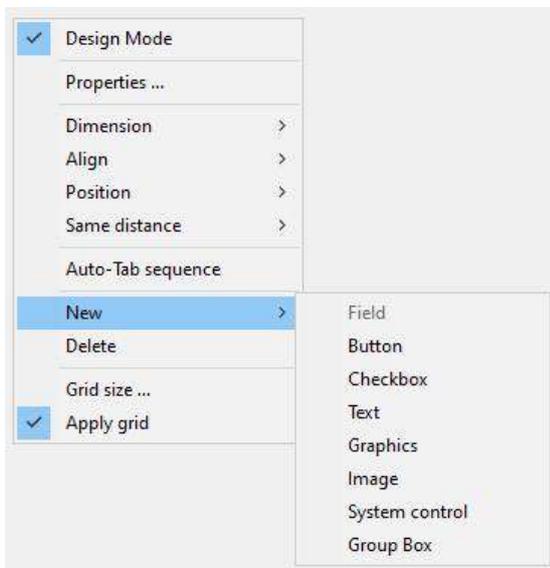
Extra custom tab for defining/scripting special features like e.g.:

- start ingest in external repository
- delete data older than 100 days
- etc.

Right click into the background to activate the **Design Mode**.

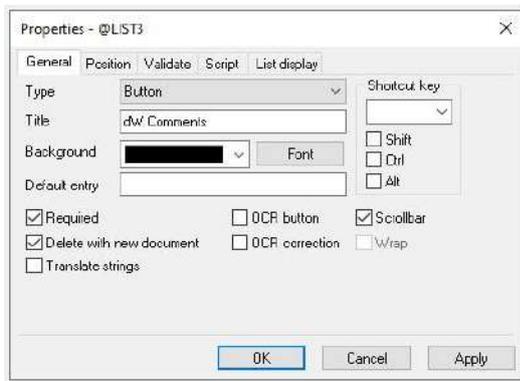


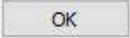
Right click to get the context menu to create check boxes, buttons, system controls and others:



Properties

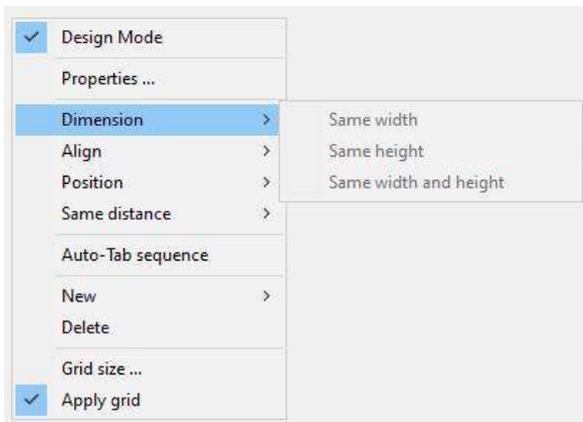
The **Properties** dialog provides a variety of tabs for making specific settings.



Here you can specify the type and label of the graphics element, as well as other attributes. Confirm and exit by pressing  .

Dimension

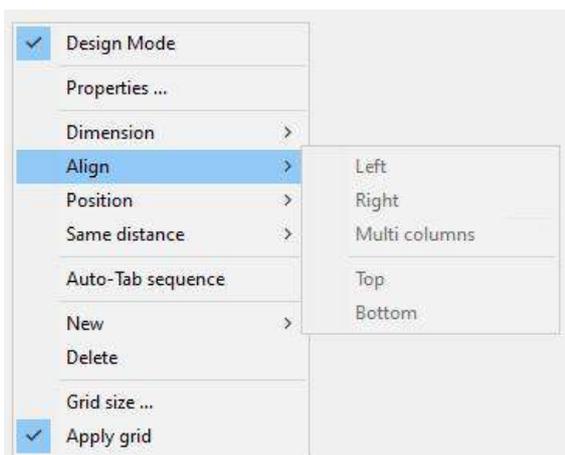
If you have selected multiple elements in the dialog box with the Shift key and the mouse, the **Dimension** function allows you to standardize the size of all the elements at the same time. Placing the mouse pointer on the **Dimension** function, another selection menu offers three commands:



With the function **Same Width**, you can scale all the selected elements to the same width. With the function **Same Height**, you can scale all the selected elements to the same height. With the function **Same Width and Height**, you can scale all the selected elements to the same width and height.

Align

If you have selected multiple elements in the dialog box with the Shift key and the mouse, the **Align** function allows you to align all the elements you have selected. Placing the mouse pointer on the **Align** function opens a selection menu beside the arrow that offers 5 commands:



The **Left** command aligns the marked elements to the left.

The **Right** aligns the marked elements to the right.

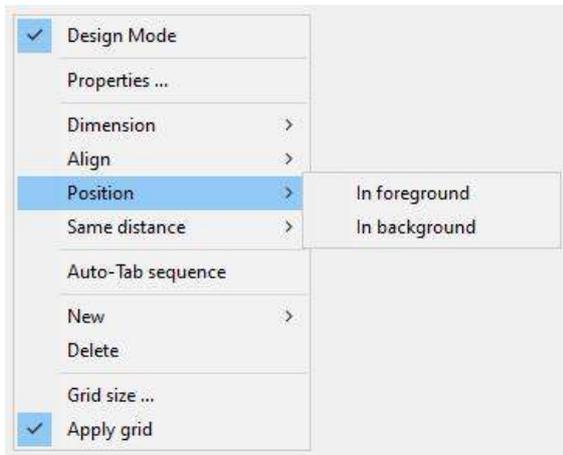
The **Multi Columns** command aligns the marked elements in multiple columns.

The **Top** command aligns the marked elements along the top.

The **Bottom** command aligns the marked elements along the bottom.

Position

Use the **Position** function to place the element you select in the foreground or background. Placing the mouse pointer on the **Position** command opens a selection menu with two commands:

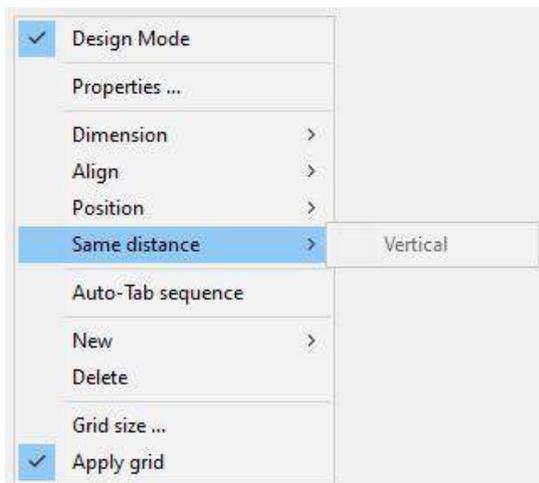


Same distance

With the **Same Distance** function, you can specify whether the vertical separation between the selected elements should be uniform.

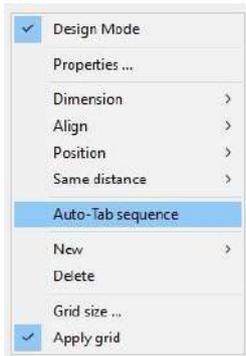
Placing the mouse pointer on the **Same Distance** function opens a selection menu offering the following command:

Vertical, means same distance in vertical dimension.



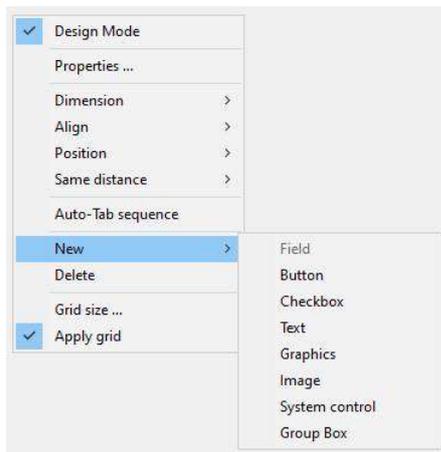
Auto Tab Sequence

You use the **Auto Tab Sequence** function to have automatically set the jump sequence for addressing the control elements when the **Tab** key is pressed. There is also the possibility to determine the order manually.



New

You can add a new element to the dialog box.



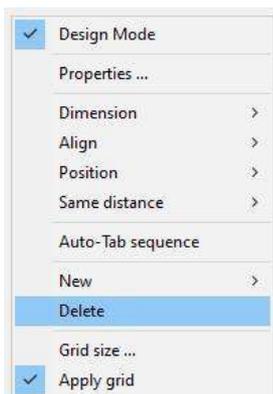
With these functions, you are able to create different elements like **Field**, **Button**, **Checkbox**, **Text**, **Graphics**, **Image** and **System Control** buttons.

Example: You use the **Graphics** button to enter graphics elements you want - backgrounds or frames - in the dialog box. These elements are for appearance only, and have no function. Click the button and place the mouse pointer where you want the graphics element to appear in the mask. Draw a frame by holding down the left mouse button and then click the frame with the right mouse button. The context menu appears.

Clicking the **Properties** function opens the **Properties** dialog box for the graphic element for example.

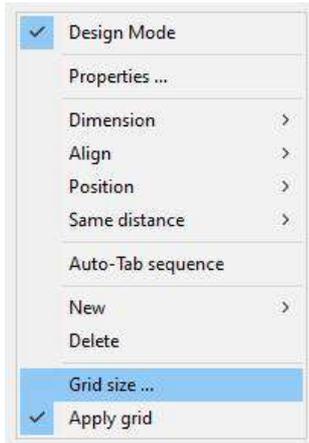
Delete

You can delete the selected element with the **Delete** function.



Grid Size

You set the size of the grid the system uses for orientation purposes with the **Grid Size...** function.



Clicking on this function opens an input mask in which you specify the desired horizontal and vertical spacing between the grid lines. Make your settings in millimeters:



Use the **Apply Grid** to turn the grid on and off. The check mark beside the menu item indicates its status.

5 Remote QA (Quality assurance)

In order to save costs, the manual checking and if necessary correction of documents can take place at arbitrary, economical places world-wide. In addition highly compressed graphic data are transferred over the Internet. The check and correction results are transferred back and processed on the production system.

The docWizz Remote QA system contains three components:

- FTP client (simple FTP client with priorities system, what to transfer first)
- Master system (on master machines, where processing is done)
- Slave system (on remote machine)

The communication between master and slave shall be done via command files, sent also using the FTP client.

RemoteQA sends all files that have not been send yet. 'Resend document' sends all files, no matter if they have been send yet or not. Origin images are never resend in case a document is in a task after Modify Pages.

Open "Error Docs" in RemoteQAManager in dWControlCenter

When holding the mouse over a document with Remote in process status (a pink "R") and you have the document loader information in Pool status file (*pool.xml.mtn) the tool tip will contain also information from loader as well.

Review status is available for normal users only on the manager side. On the loader it is not available for normal users to prevent accidental document sending on the manager side.

The error and reject statuses were made to not be available for normal users. Usually documents reach this status automatically by a service, therefore normal users should not be able to change the document status to these ones. The review status has a special meaning on the loader side, because it sends back documents to the manager, therefore, this operation should be done by users with more privileges.

See the manual [docWizz ReferenceBook](#) for details.

6 Backup, Autosave, Update

Auto-Update

If an automatic update is available you will get a message:

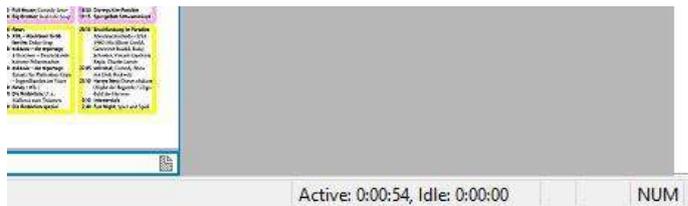
Here you select when the system shall remind you again to close docWizz and reopen it again. While reopening the update process is done.

Auto-Save

In case of system errors the auto-save functionality is very useful to save already done work.

Auto-save is done:

- every 10 minutes of inactivity (Idle status)
- every 30 minutes when working (Active status), a short message will be shown



- Auto-save files are stored in Pool folder additionally to the document files
- If docWizz crashes and will be restarted you can select if you want to go back to the auto-saved status or not.
- Auto-save file will be deleted if the document is the next job
- Auto-save works in all jobs except Exported task
- On regular close of documents or docWizz, auto-saves are deleted
- <docID>AS<timestamp>.xml is created by rename after <docID>AS<timestamp>.zip is successfully stored.
- In ScanClient all page based data is stored immediately on disk. So no auto-save needed.
- A message will be shown when opening a document that was auto-saved

Backup and remove

Behavior is like this:

- before backup, document is set to reduced state to make document smaller
- Backup & Remove tells to restore a document always to Review structure and text (no need to have it in exported, because if the user wants to restore than he wants to work on it)
- you need to select "restore pool data" before you may route the document.

Copyright © 2022 CCS Content Conversion Specialists GmbH. All rights reserved.

No part of this publication may be reproduced, stored in databases, or transferred in any form (electronically, photo-mechanically, chemically, manually, or otherwise) without the express written permission of CCS Content Conversion Specialists GmbH. The software described in this manual is licensed software that may be used only in compliance with the licensing terms and conditions. CCS GmbH reserves the right to make changes to the content of this manual without notice. CCS GmbH makes no guarantee regarding the accuracy of the information provided in this manual. Microsoft, and Windows are registered trademarks of the Microsoft Corporation.

Product or company names that are mentioned may be trademarks or registered trademarks of the respective company. CCS GmbH uses these names and trademarks in the following manual merely for explanatory purposes and for the benefit of the respective user, and such use does not imply trademark infringement.

Under this software license, you are only permitted to reproduce materials that are not protected by copyright laws. This excludes only materials where you hold the copyright and/or legal permission to reproduce copyrighted materials. If you are uncertain about the copyright status of certain materials then please seek legal counsel. CCS GmbH holds no liability over copyright violations resulting from the use of this software.

Last updated: 17.03.2022

CCS Content Conversion Specialists GmbH

Weidestraße 134

22083 Hamburg, Germany

Phone: +49-(0)40-228582990

E-Mail: info@content-conversion.com

Website: www.content-conversion.com