Universal Capturing Client - UCC

Operating manual for the Universal Capturing Client

Revision	Date	Version	Amendments
1	01.06.2012	1.0	First draft
2	27.04.2013	1.1	Update for version 1.1.
3	18.05.2014	1.2	Revision for version 1.2
4	24.05.2014	1.2	Minor textual changes and new screenshots
5	29.10.2014	2.0	Adjustments for version 2.0 and new screenshots

Contents

1 What is the Universal Capturing Client?

The Universal Capturing Client is a scanner management application specially designed for the era of mass digitisation. The single UCC interface can be used with different devices. This means that digitisation staff can operate a range of scanners using just one application, thus avoiding the need for time-consuming familiarisation.

The UCC is designed for touchscreen operation. It contains no overly complex functions that are not required for mass digitisation scanning. Equally, there are no complicated key combinations or unnecessary dialogues with complex settings.

Please note that administrator access is required when installing the UCC. Following installation, the settings can then be configured by the same users who operate the scanners.

2 Overview of program interface

The following screen will appear when the program is opened.

Figure 1: UCC start page

The UCC program interface is very simple and clearly structured. Consecutive workflow steps are performed from left to right. The whole workflow is roughly divided into the following steps.

Step 1 involves creating new projects, opening existing projects or importing them directly from Goobi. Step 2 involves setting the scan parameters that determine the areas to be scanned. Step 3

covers the actual scanning process and the simultaneous capture of structure elements, metadata and pagination data. Step 4 provides an opportunity to correct the results of automatic text recognition. Finally, step 5 involves publishing your results, e.g. as a PDF file, as an export to the intranda viewer or as a Goobi export.

Each of the above steps is explained in detail in this document.

3 Basic settings

Before you can use the UCC to scan your material, you will need to enter a few basic settings in several areas, e.g. for the scanner, the configuration of structure data and metadata and communication with Goobi.

First of all, open the settings dialogue entitled `Preferences`. To do this, select `Program settings` from the `File` menu. You can now choose between various types of settings in the left sidebar.

3.1 General

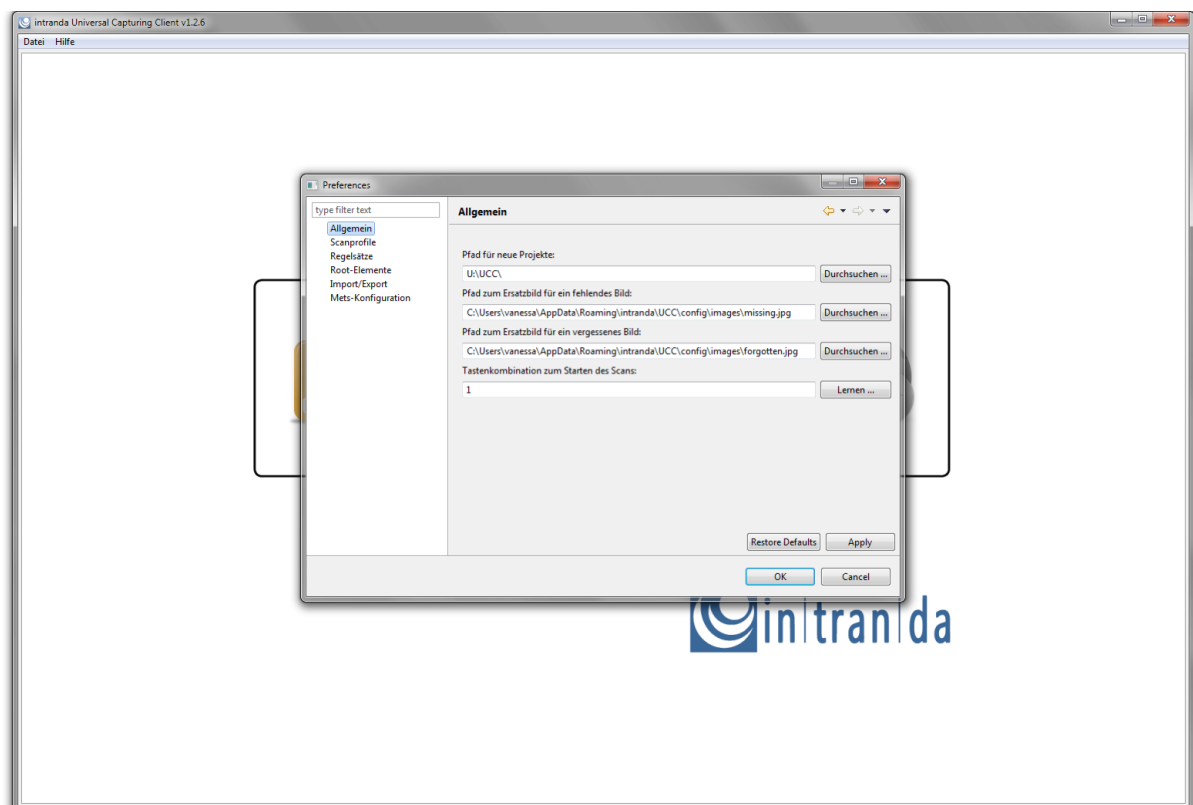Under the heading `General` you can specify various basic settings for the entire program.

Figure 2: Choosing your general settings

These settings do not usually need to be modified individually, although after installing the UCC it is important to check the value entered in the field `Path for new projects`. We recommend that you specify a path with sufficient storage capacity for your planned digitisation projects.

Digitised material generally requires a lot of storage capacity, so it could be a good idea to use a dedicated partition or hard disk.

3.2   Scan profiles

Under the heading `Scan profiles` you will need to enter your basic hardware operation settings. Depending on your configuration, this screen will display a number of scanners that have been entered together with their respective parameters. One of the options available at this point is to view the `StorageScanner`, which functions as a virtual scanner and therefore allows you to scan directly out of the file system. This makes it possible to edit existing image files.

Figure 3: Configuration of available scanners

To edit a particular scanner, simply choose it from the list and click on the `Edit` button underneath the list. You can then specify the individual properties of that scanner in the dialogue box and assign it a memorable name. You can also add new scanners. Just click on the `Add` button. All the scanners configured in this area will then be available for individual scanning projects using the UCC.

3.3   Rulesets

Under the heading `Rulesets` within the settings menu you can configure individual and document type-specific rulesets. These rulesets are control files that specify what structure data and metadata elements should be available when working with the UCC. This allows you to stipulate precisely how you want your staff to capture data on the material being scanned.
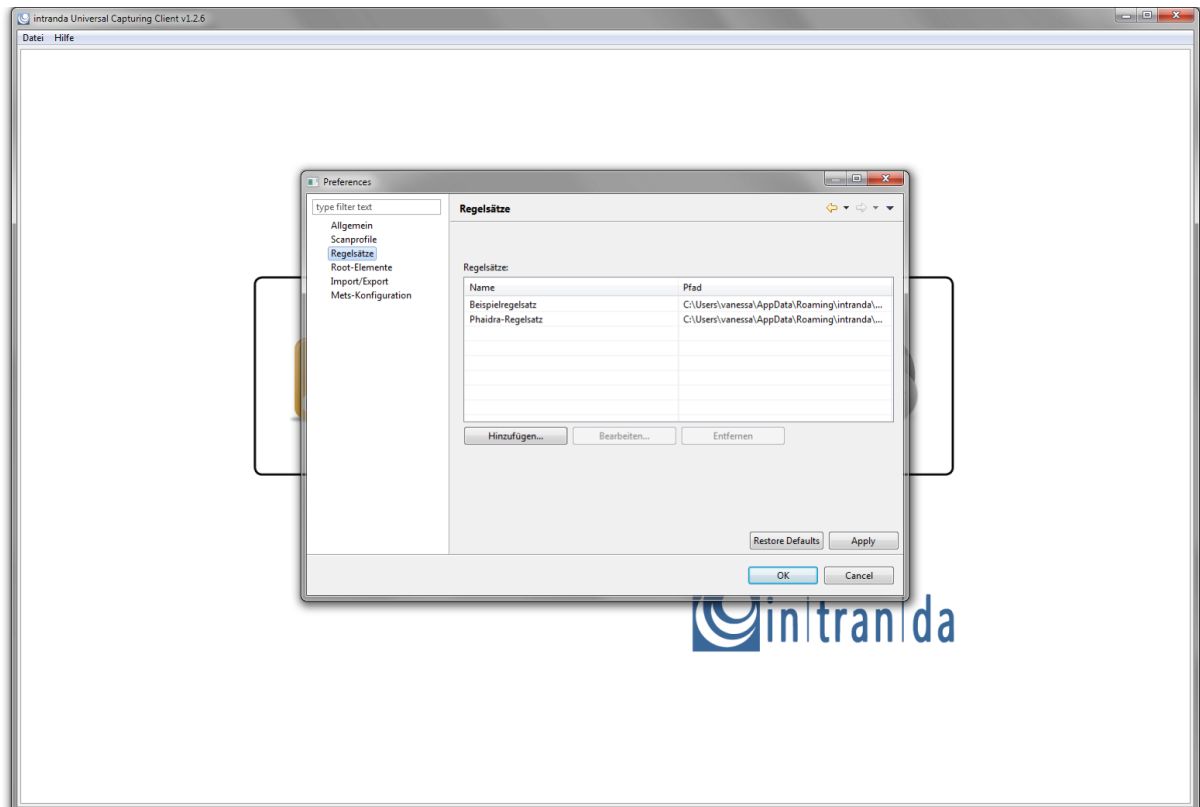
Figure 4: Specifying rulesets

The dialogue box allows you to specify individual rulesets for use by the UCC. They will then be available for your scanning projects.

Important note

If you use the UCC in conjunction with the workflow management application Goobi, there is no need to configure this item. The software tools are designed to work together. Goobi automatically assigns the correct ruleset to the UCC, which then uses it for the designated scanning project independently of the configuration specified under Preferences.

3.4 Root elements

Within the UCC, you can also manage the way your staff work with rulesets and capture data on the scanned material under the heading `Root elements`.

Figure 5: Specifying the available main elements as document types

This configuration dialogue allows you specify which publication types you wish to offer within the UCC when creating new projects.

Important note

Once again, if you use the UCC together with Goobi, there is no need to configure this item.

3.5 Import / Export

The ultimate goal of working with the UCC on scanning projects is to export the resulting data. The application gives you a range of options for doing so. These can be configured under the heading `Import/Export`.

Figure 6: Configuring the import and export interfaces

To configure an export type, first select an element from the list and then click the `Edit` button to make any changes you require.
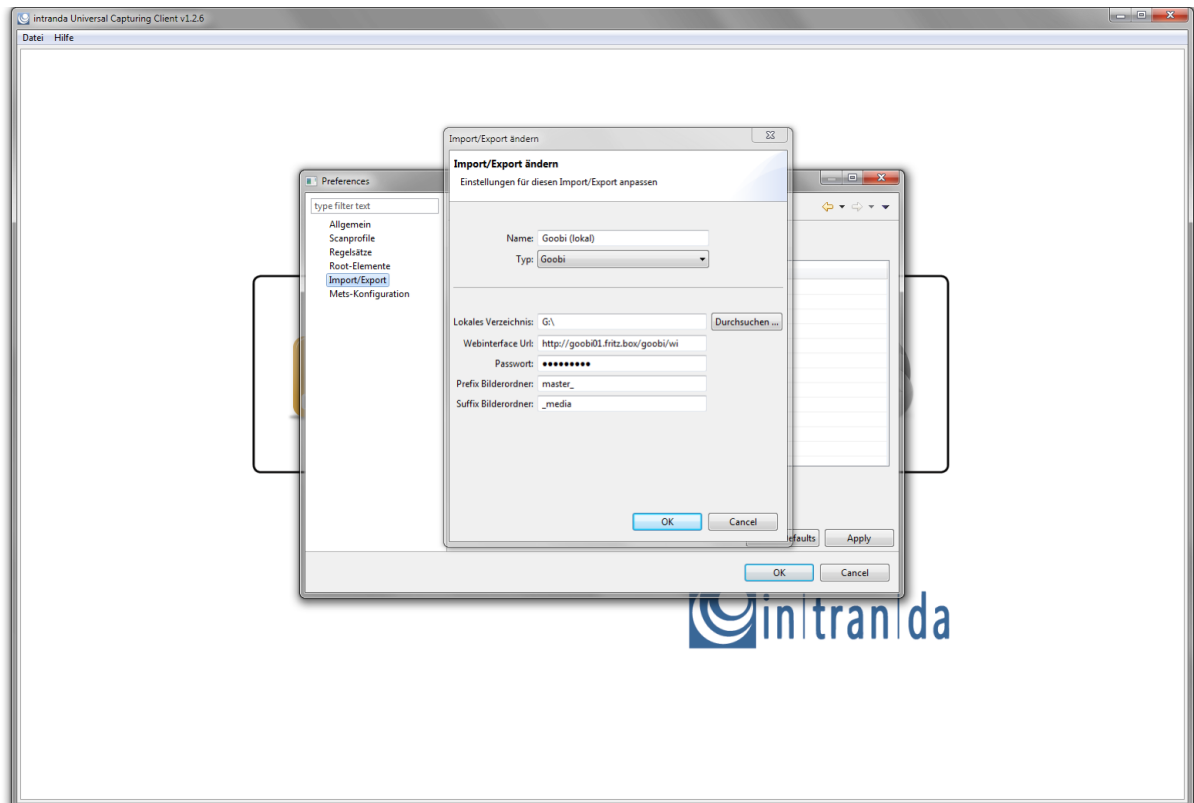
Figure 7: Configuring the interface to Goobi

When you configure your system to work with Goobi, you will need to correctly specify a number of parameters to ensure that both programs are able to communicate. As well as entering the correct Goobi URL, you must particularly ensure that the `Password` for the Web API interface is set correctly and that the folder names are configured in exactly the same way (`Prefix` and `Suffix`) as in Goobi. These settings must be correct for data to be imported and exported smoothly between Goobi and the UCC.

3.6 METS configuration

If you wish to use the UCC to publish your digitised material directly together with your metadata (e.g. by exporting the data to the intranda viewer), you can use the METS configuration dialogue box to specify the content that you want to appear in certain metadata fields. You can, for example, specify details relating to the owner, contact addresses and other similar settings.
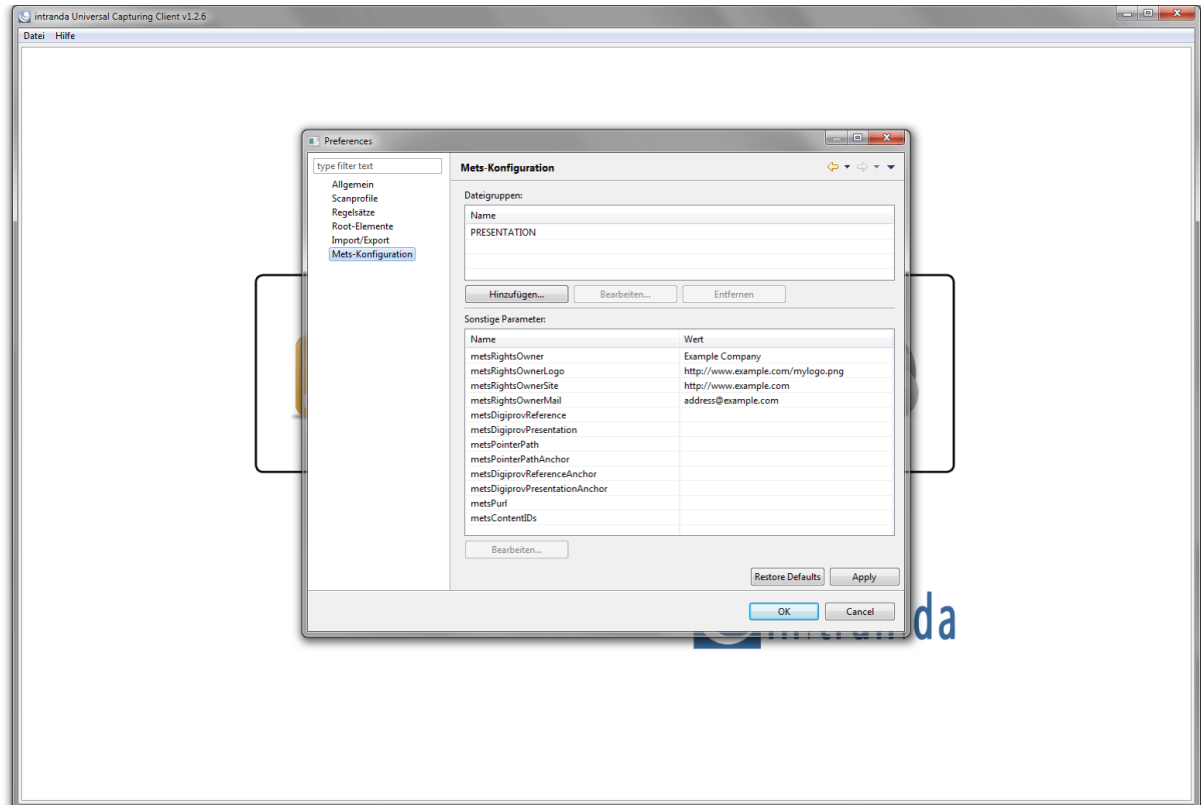

Figure 8: METS configuration settings

Once again, simply choose those items from the list that you wish to edit, click the Edit button and make whatever changes are required.

4 Working with the UCC

As a general rule, there are five steps involved in working with the UCC. A detailed description of each step is given below. The workflow steps are performed in the order displayed on the application start page (i.e. from left to right) simply by clicking the appropriate icon. This will open the dialogue window for that step in order to begin processing.

Figure 9: UCC start page with five icons for the individual workflow steps

When you start the program, all the icons referring to subsequent workflow steps will be deactivated. First you need to create a new project or open an existing one.

4.1 Creating a project

Before you can start working with the UCC, you will need to open an existing project or create a new one. To do this, select the first icon on the start page.

Figure 10: Opening or creating a project

This will open the dialogue box entitled `Project settings`, where you can create a new project or open an existing one. You can also import an existing project from Goobi.
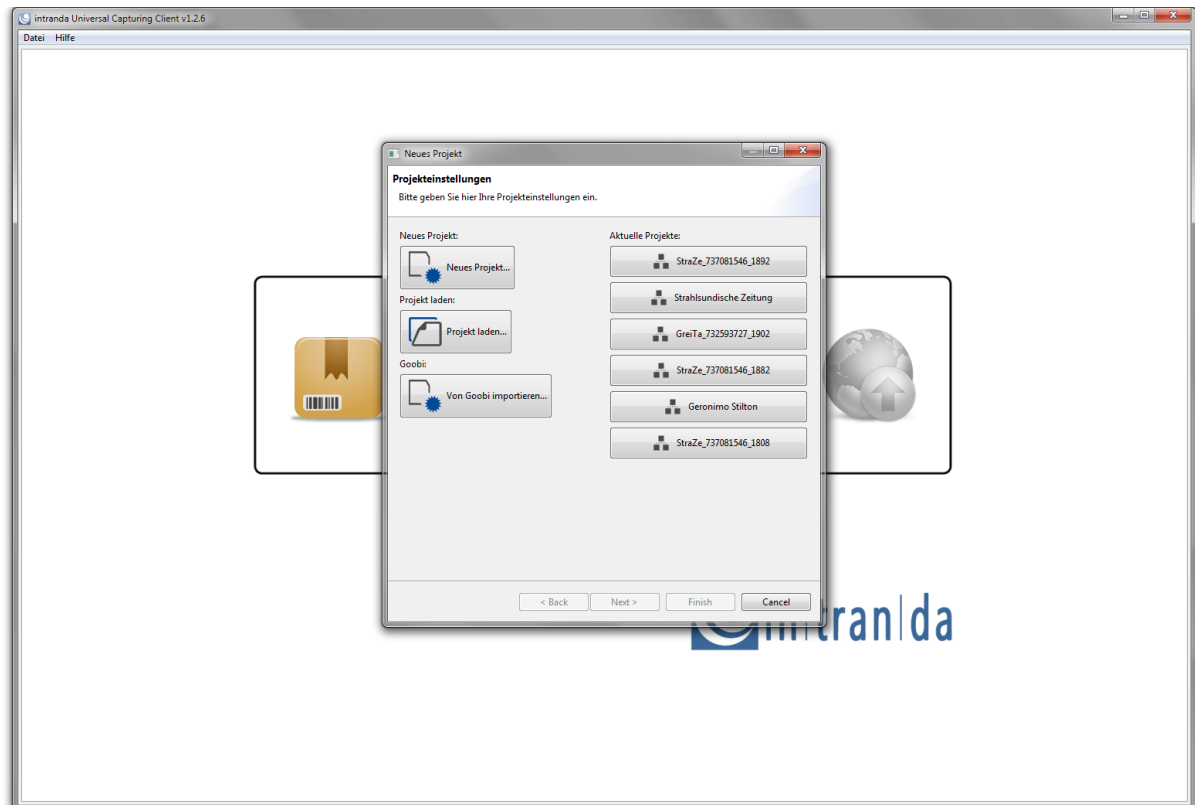
Figure 11: Project settings

As well as the option to create a new project, you can reopen an existing one. A list of the most recent UCC projects is displayed on the right of the box. You can either select the project you require from this list or load any other project from the file system.

4.1.1 Importing a project from Goobi

If you are using the UCC together with Goobi and the interface between the two applications has been fully and correctly configured, you can now import data from Goobi. However, before doing so, you must first ensure that you have both read and write access to your Goobi work directory. This will need to be set up on your system (e.g. using the intranda Mount Tool) as a network drive (usually `G:`). Next, log in to Goobi and accept a task (e.g. scanning) that has been allocated to you. This will instruct Goobi to provide you with a directory for that task within your work directory.

Important note

The UCC can only import data from Goobi if the configuration is correct, the Goobi work directory can be accessed and a workflow step has been accepted in Goobi.

In the project settings dialogue box, select the button `Import from Goobi`. This will open another dialogue window in which you can choose the required folder. Here you will need to select the folder provided by Goobi within your work directory (e.g. `G:\MyVolume_12345\`).

As soon as you have confirmed the import path, the UCC will import the data from Goobi, and you can begin the work of scanning and indexing your material and capturing data.

4.1.2 Creating a project manually

If you are not working with Goobi as well as the UCC, you can select the `New project` button in the `Project settings` window. This will open a new dialogue in which you can specify a name for the project.
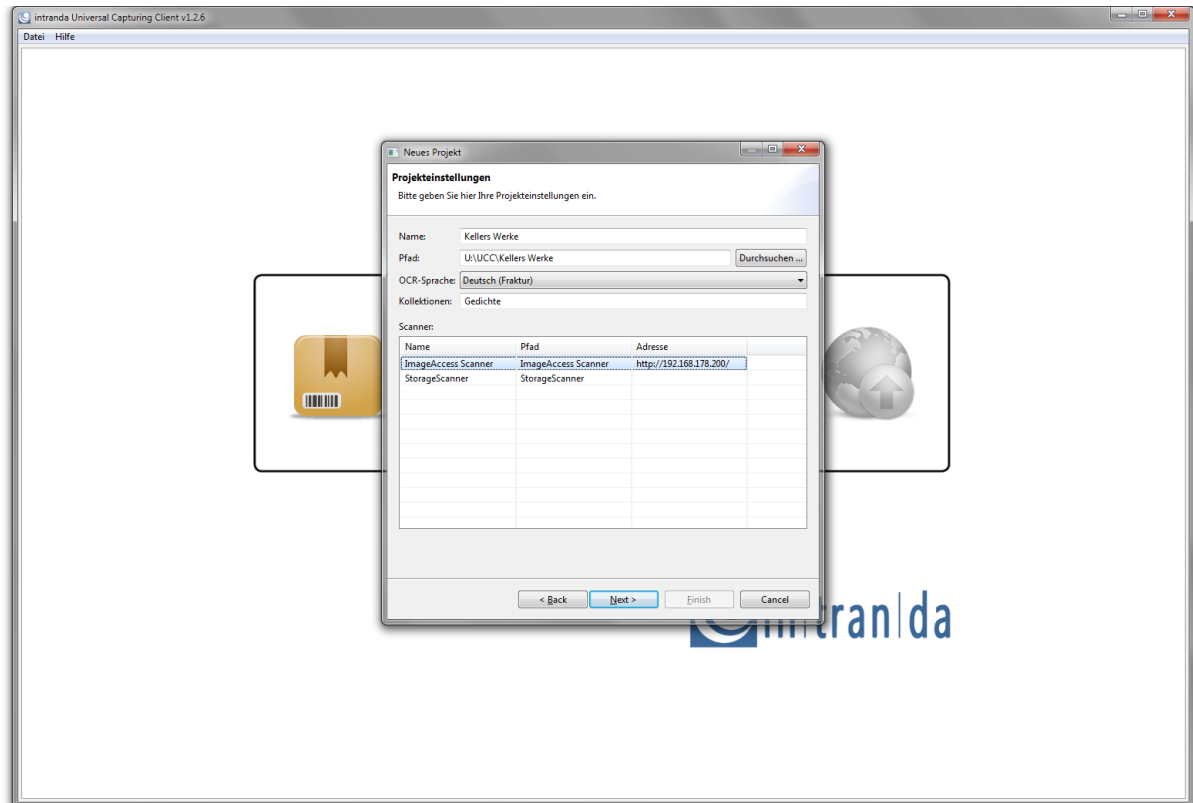

Figure 12: Specifying the project settings

Next, specify the typeface used in the book in combination with the language in order to configure subsequent text recognition for the source work. To conclude, select the scanner to be used for the current project and click the `Next` button.

In the following `Scanner settings` dialogue, you can again define various settings to be applied to the current scanning project, e.g. whether a glass plate is to be used, whether the book is to be scanned in a rotated position and whether the images are to be cropped using certain parameters.
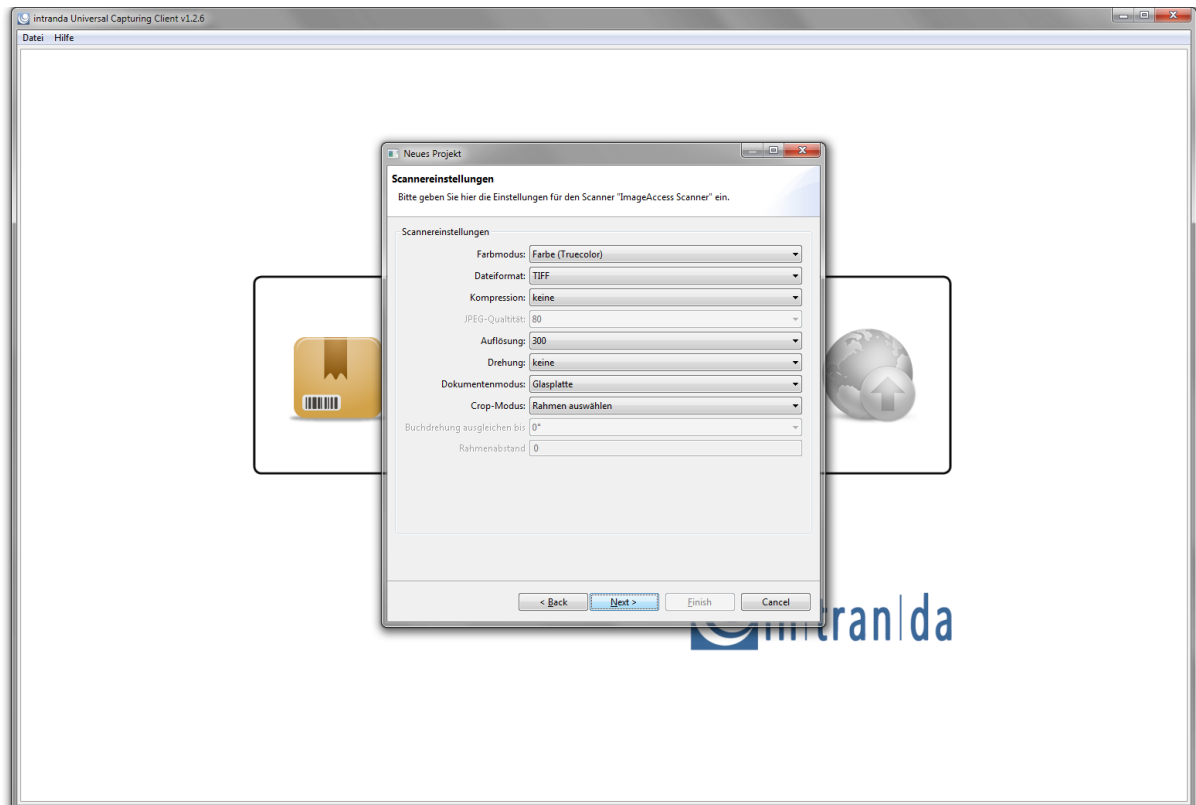
Figure 13: Configuring the scanner settings

Click the `Next` button to open the `Ruleset` dialogue. Here you need to select the required ruleset and specify the document type for the current scanning project.
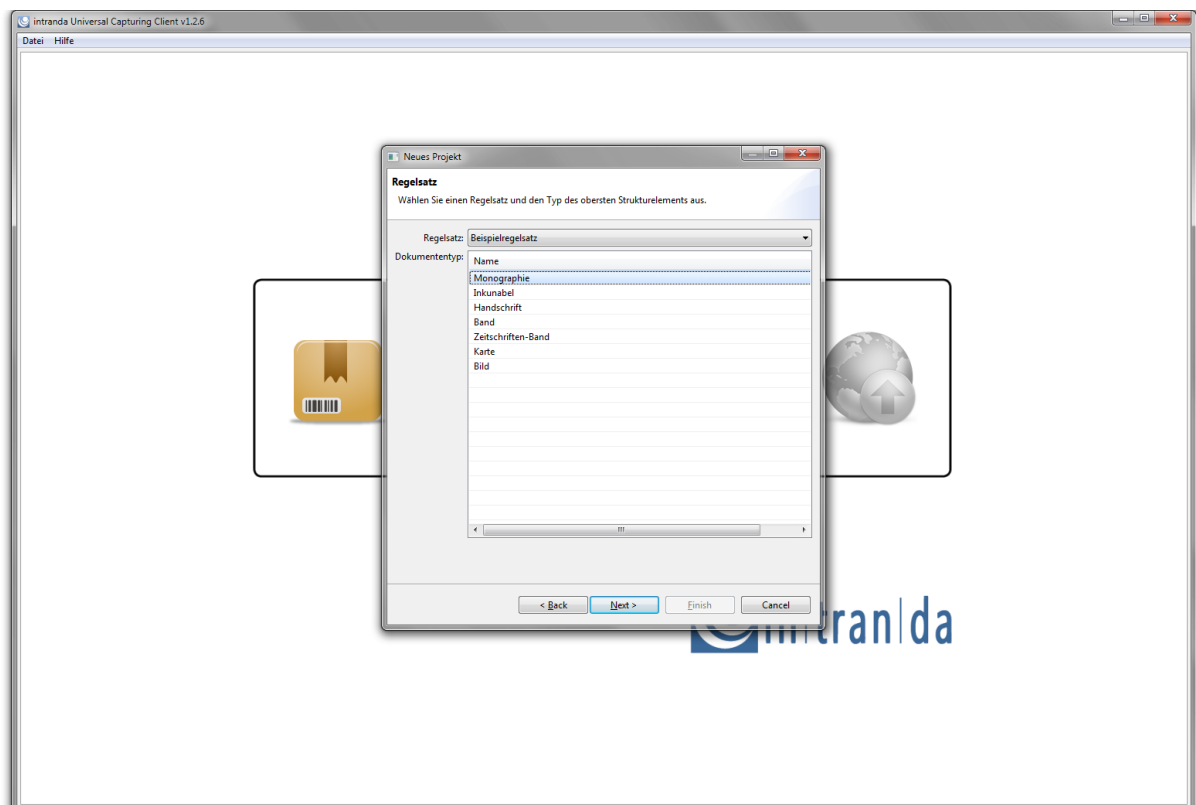

Figure 14: Selecting the required ruleset and document type

In the next and final dialogue window entitled `Metadata` you can enter a range of basic metadata. Please note that fields marked with an asterisk are mandatory and must therefore be completed. If you wish to capture additional metadata, you can do so by selecting either of the buttons `Add new metadatum` or `New person`.
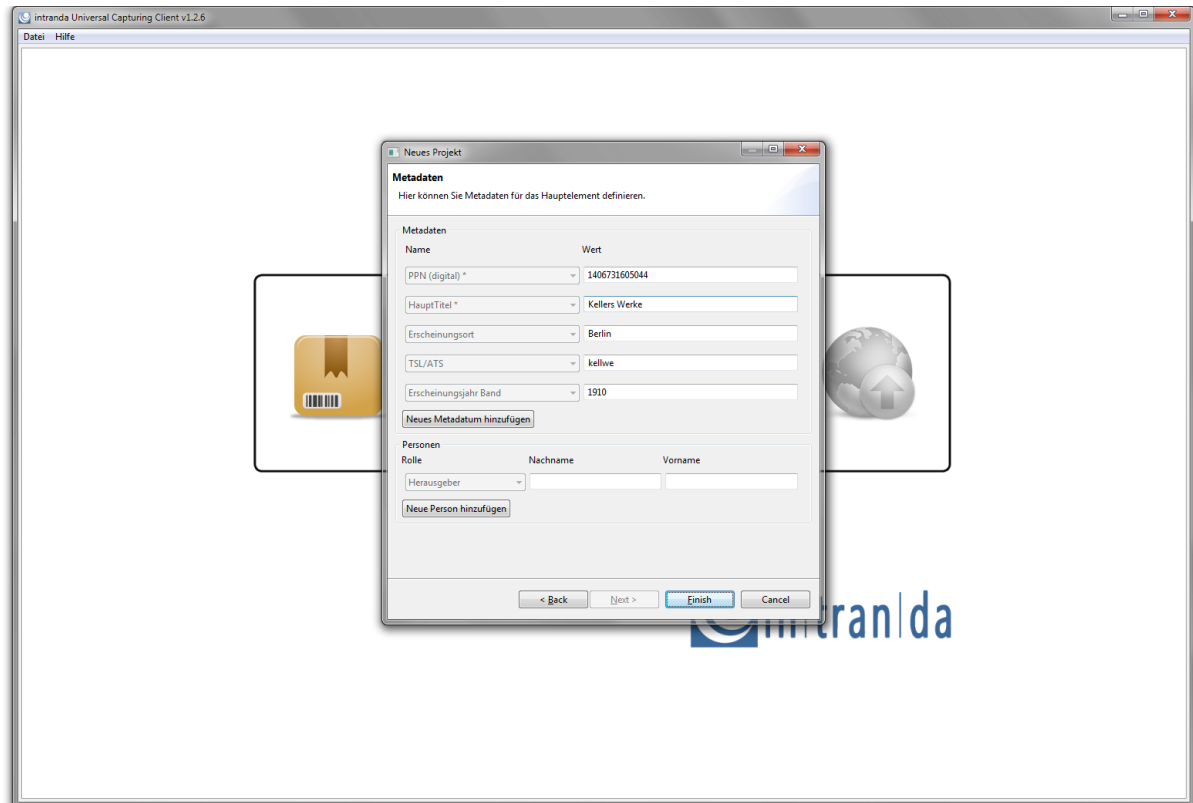

Figure 15: Metadata

Once all the fields have been completed, select the `Finish` button to create the project.

4.1.3 Opening an existing project

If you subsequently wish to reopen an existing project, you can do so directly from the `Project settings` dialogue box. Rather than creating a new project manually or importing a project from Goobi, you can also choose to reopen one from the list of most recent UCC projects shown on the right of the dialogue box. If the project you require is no longer listed, you can locate it by clicking the `Load project` button. This will open a dialogue that allows you to open the UCC project from your file system.

Once you have loaded the project, it will be opened by the UCC and will then be available for further processing.

4.2 Parameter setting

Once you have opened a project, you can proceed with the next workflow step. This involves setting the parameters. The corresponding icon will now be activated on the start page.

Figure 16: Start page with the option to perform the next activated workflow step (parameter setting)

Click the second icon on the start page to open the parameter settings view.
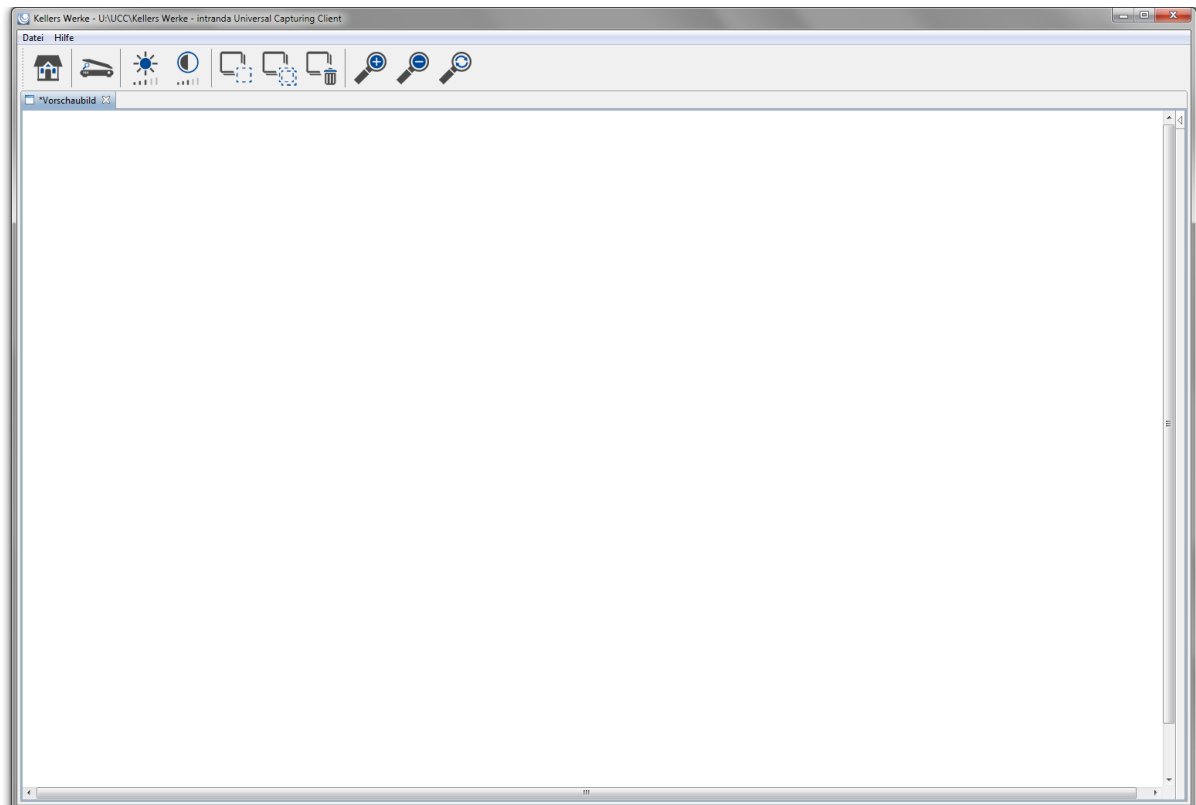
Figure 17: Parameter settings dialogue

In the UCC's parameter settings view you will find a menu bar at the top of the window. First click the `Scan` icon to request a preview scan from the scanner. This will produce a scanned image that you can use to set your image parameters.
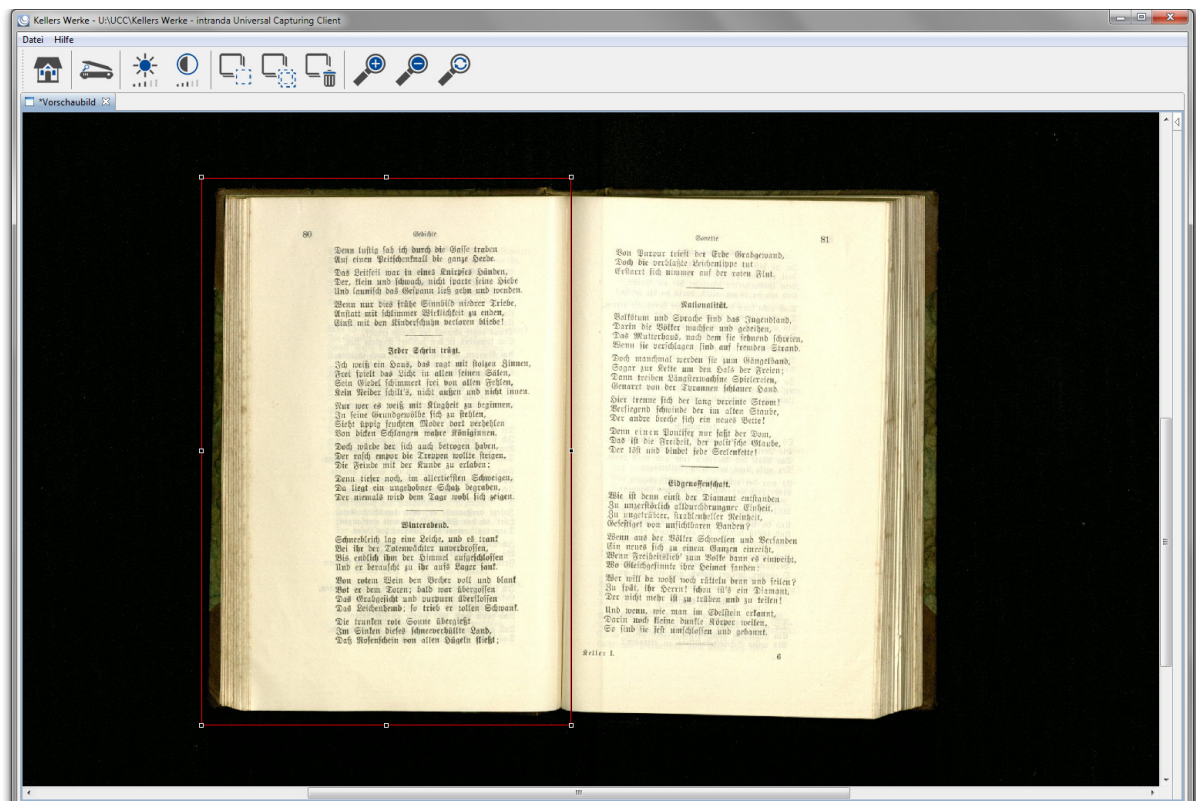
Figure 18: Left image parameter setting

To set the image parameters, click the `Simple border` icon and create a border around the scanned left page of the book. The border will appear in red. The size and position can still be changed. Make sure that your frame covers the entire image and ideally some additional space around the text. We recommend that you place the border beyond the centre spine to ensure that you always scan the full image.
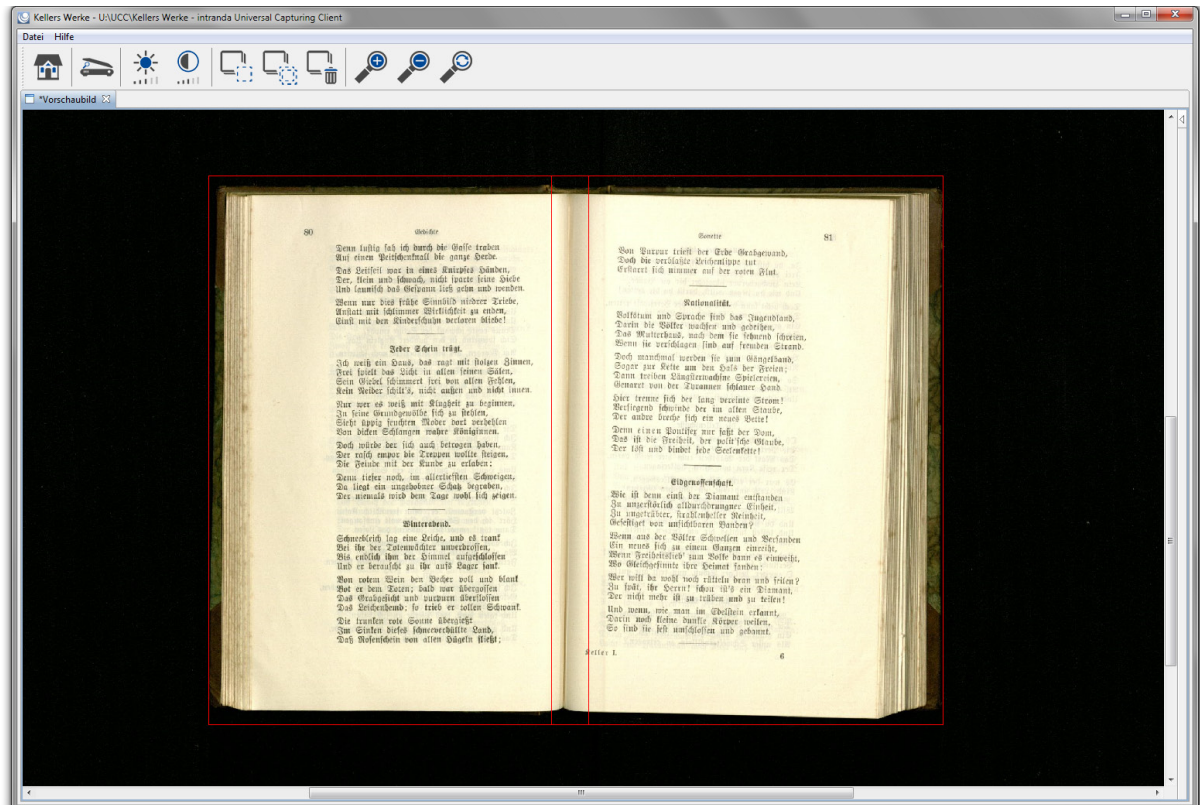

Figure 19: Setting both left-page and right-page borders

Once you have set the left-page borders, simply click the `Duplicate` icon. This will automatically copy the borders used on the left to the right with an overlap in the centre.

Borders can be adjusted individually or together. By pressing the shift key, you can highlight several borders and move them together. You can also use the keyboard to reposition borders. Borders that are either incorrect or no longer wanted can be removed at any time by clicking the `Delete` icon.

Once you have set the scanning parameters, click the `Start page` icon to return to the project overview.

4.3 Scanning

When you have completed all the above preparation, you can now begin to use the UCC for its intended purpose. If required, this can involve a range of tasks. As well as the actual job of scanning, it can be used to capture structure and pagination data for your digitised material.

Figure 20: Start page showing activated scanner icon

To commence scanning, simply click the `Scan` icon. This will open the UCC's scanner interface, which contains various sections for continued processing.
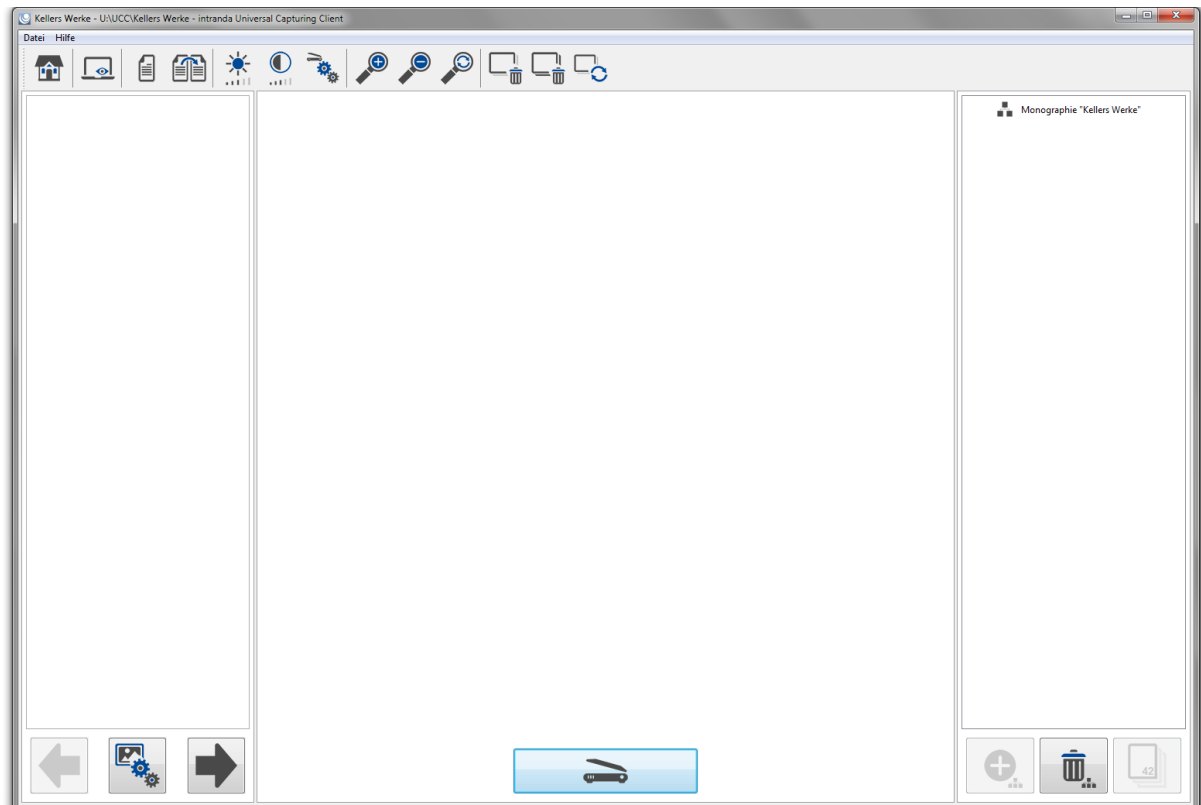
Figure 21: The UCC's scanner interface

At the top of the window you will find a series of icons that are used to call different functions. You can identify each function simply by holding the mouse briefly over the icon.

Icon	Description
Start page	This will take you back at any time to the project overview page.
Scan preview	This opens the window for setting your scanning parameters (borders).
Single page	If you select this icon, the UCC will display only single pages in the scanner interface.
Double page	If you select this icon, the UCC will display double single pages in the scanner interface, i.e. in the same way as the actual open book.
Swap page order	You can select this icon to specify whether the book begins with a right or left page. If the book cover is also to be scanned, the book will normally start on the right.
Bitonal value (brightness)	Use this icon to specify a bitonal value for scanning purposes. This setting will only be applied to the next scan.
Contrast	This icon can be used to specify the contrast level when scanning. The setting will only be applied to the next scan.

Zoom	The zoom icons allow you to specify how large the scans should be displayed. This function can be controlled using the mouse wheel as well as the menu items.
Delete last border	Select this icon to delete the last OCR border you set.
Delete all borders	Select this icon to delete all the current OCR borders.
Restore all borders	Select this icon to restore all the borders used for the last structuring operation.
Omit pagination	This icon is used to specify whether pagination is to be omitted on the left or right page.

Other functions can be called using the icons located at the bottom of the window on the left and right:

Icon	Description
Previous page	This icon allows you to navigate back through the sequence of images to the previous page.
Menu	Select this icon to open the context menu for the selected images. You can also open this menu by clicking the right mouse over a preview image.
Next page	This icon allows you to navigate through the sequence of images to the following page.
Scan	Instructs the UCC to perform a scan.
Structure element	Add or amend an existing structure element.
Pagination	Specify the pagination for the current page and (if required) subsequent pages.

4.3.1 Scanning

You can instruct the UCC to perform a scan by clicking the `Scan` icon. Alternatively, you can do so using the scanner's foot pedal or the glass plate. To do this, you will first need to activate the checkbox `Scan in batch mode`.
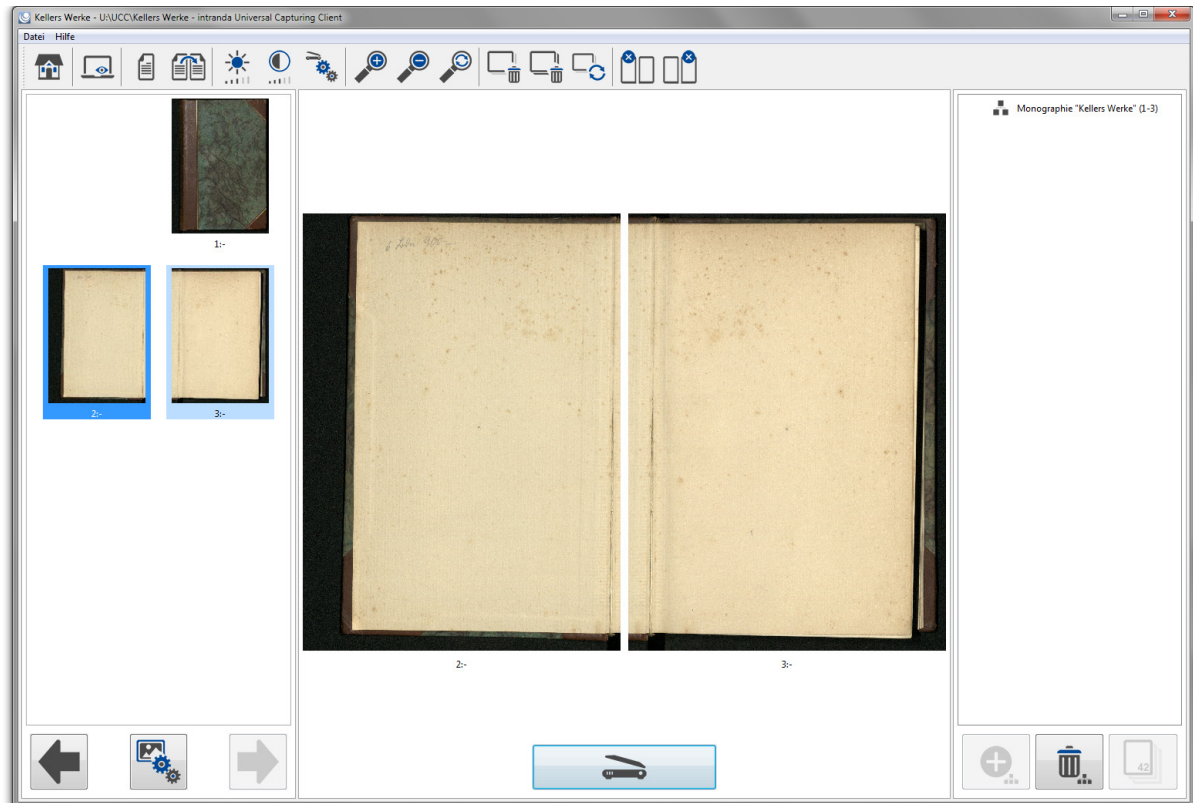

Figure 22: Scanner interface after performing several scans

Following each scan, the UCC will display the number of images you specified when you set the scan parameters (borders). The resulting preview images are shown on the left sidebar, where they can be edited, moved, cut/copied/pasted or deleted.

4.3.2 Capturing structure data

As well as scanning your source works, the UCC allows you to capture the corresponding structure data. Simply click first on the page to which you want to assign a structure element. In addition to classifying the page, you may also wish to assign one or more metadata to the new structure element. To do this, use the mouse again to draw one or more small borders that will then be used in the UCC's integrated text recognition tool. For example, you might choose to draw a border around a page heading.
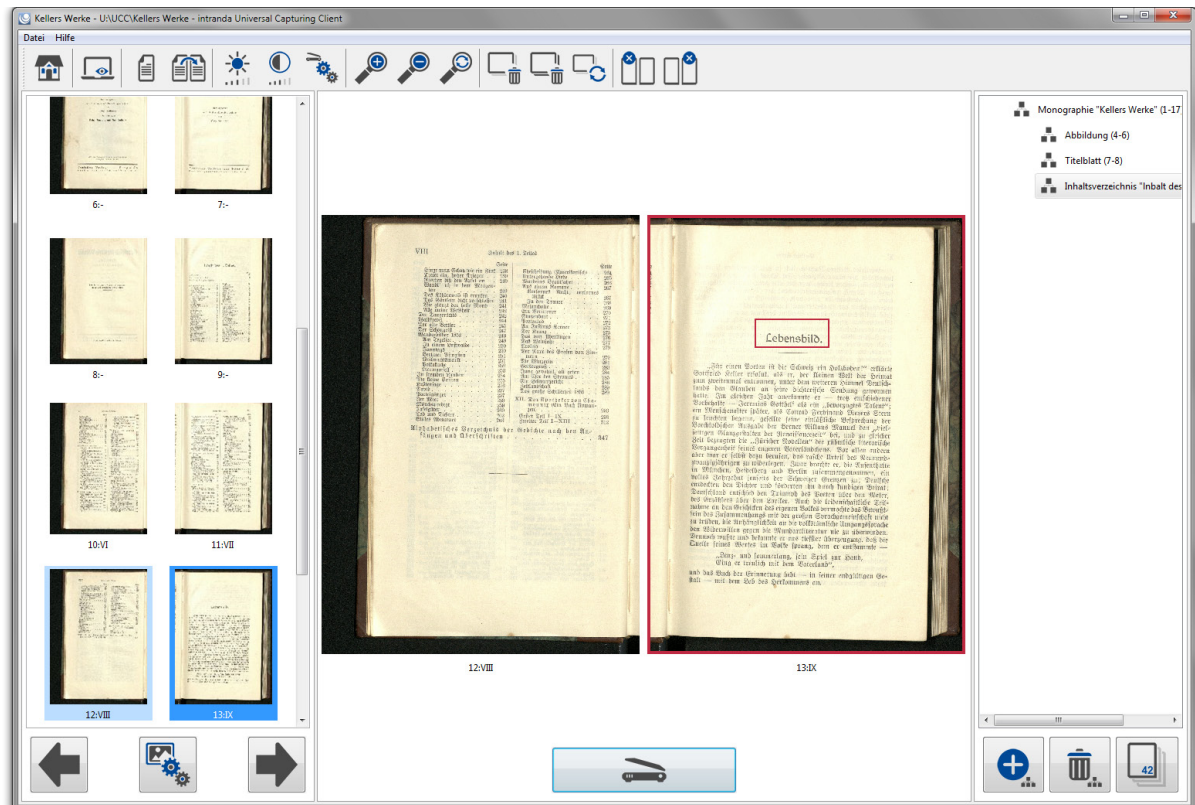
Figure 23: Drawing a border around a heading

Once you have highlighted a page or drawn a border, click the structure element icon in the bottom right corner. This will open a new dialogue box in which you can specify a type for the chosen structure element and where it should be located within the structure hierarchy.
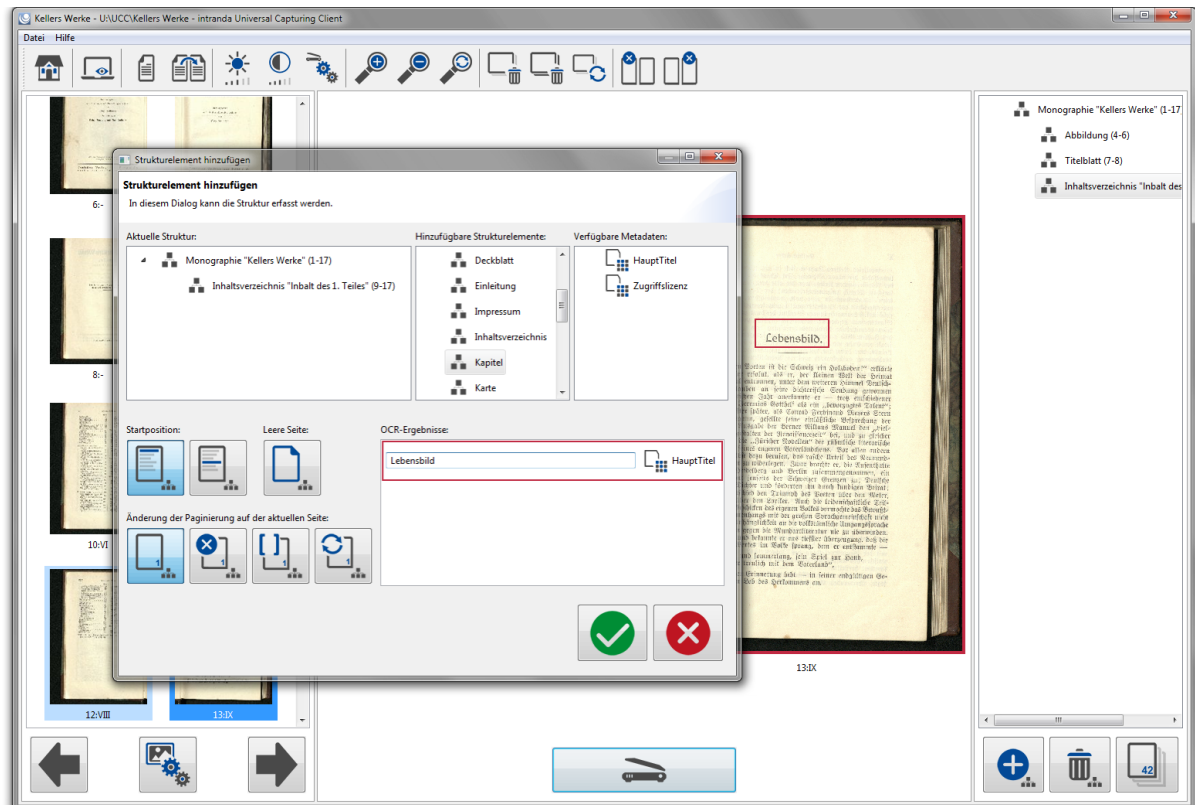
Figure 24: Adding a new structure element

In this dialogue box you can select the existing element within which you want to generate the new structure element as a sub-element. In the centre of the dialogue you can specify a type for the new structure element. The automatically generated OCR results for the area around which you placed a border will be displayed in the bottom right corner of this window (usually as the main title for the new structure element).

At the bottom of the dialogue box, you can also specify whether the structure element is the only one on this page or whether the previous structure element is also located on the current page (e.g. the end of the preceding chapter).

If you want to assign the OCR result to another metadatum instead of the main title, you can select that metadatum from the list in the top right corner of the dialogue and simply drag it down to the OCR'd text element.

To conclude, click the `Apply` button to save the data and add the newly generated structure element to the overall list of the source work's structure elements in the right-hand sidebar.

4.3.3 Simultaneous capture of multiple structure data

The UCC allows you to capture various structure data at the same time. To do this, simply draw a number of borders on the required page one after the other.
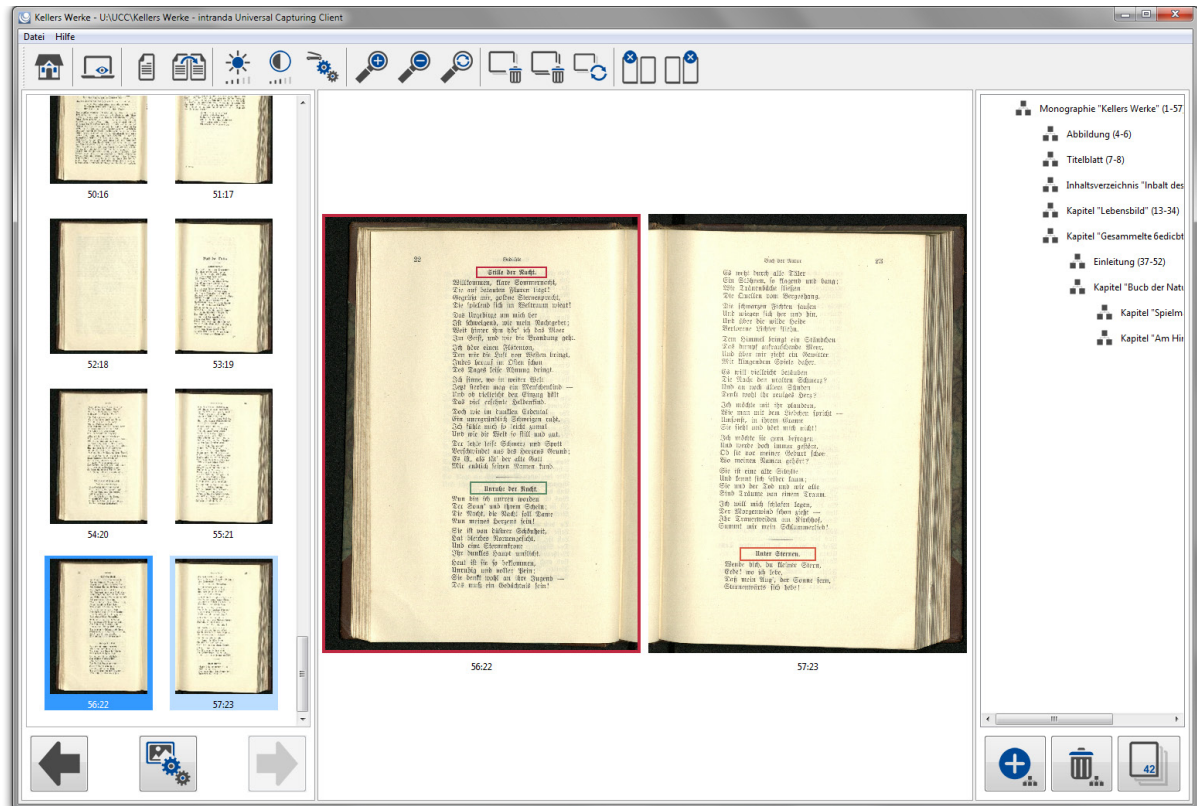

Figure 25: Placing several borders in order to generate multiple structure elements at the same time

Next, select the `Structure data` icon to open the corresponding dialogue, where you will find the OCR results for each border you have drawn. Choose the structure element within which the new elements are to be assigned, and then select the structure element type as described above. In this case, the OCR texts from inside your borders will be automatically assigned to the available metadata. By way of example, you can assign both the main title and a subtitle to a single structure element.
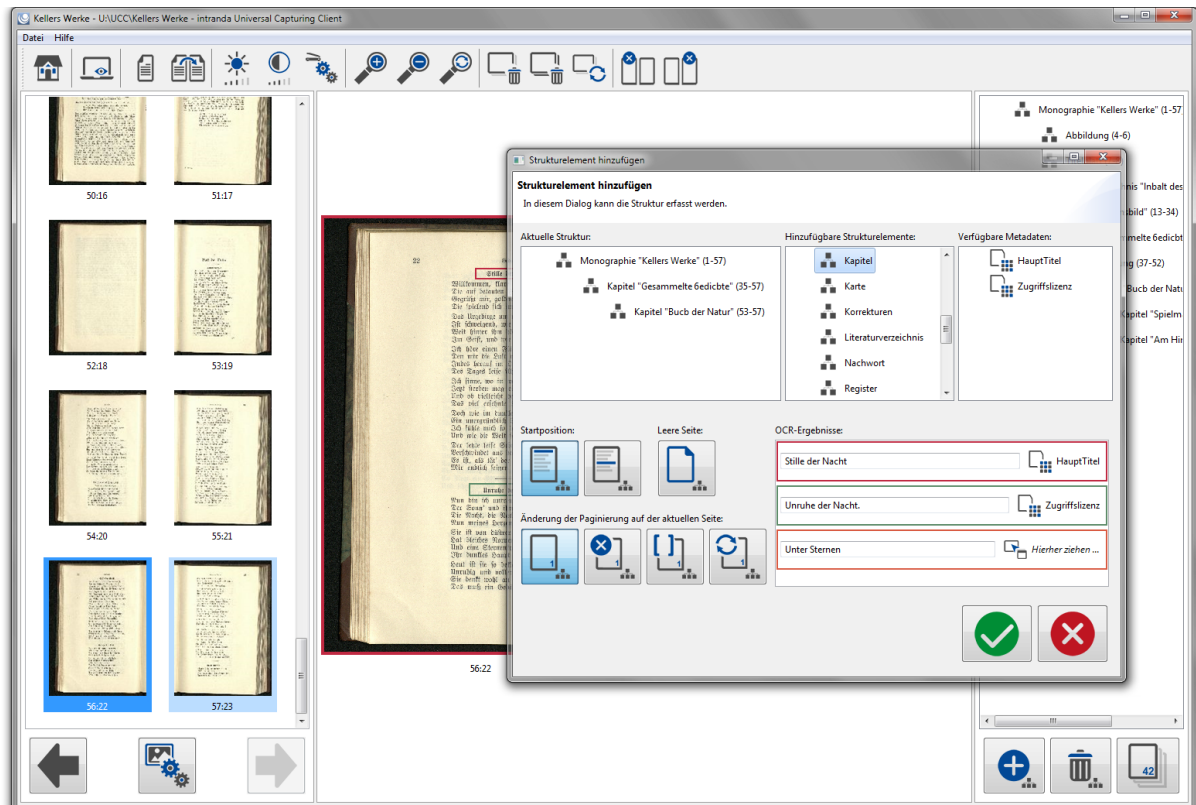
Figure 26: Assigning several metadata to a single structure element

Alternatively, you can generate separate structure elements for each item surrounded by a border. First select the structure element to which they are to be assigned as sub-elements, and then simply double click the structure element type to generate a separate structure element for each bordered item you have selected. The content of the bordered item will then be assigned to each element as the main title.
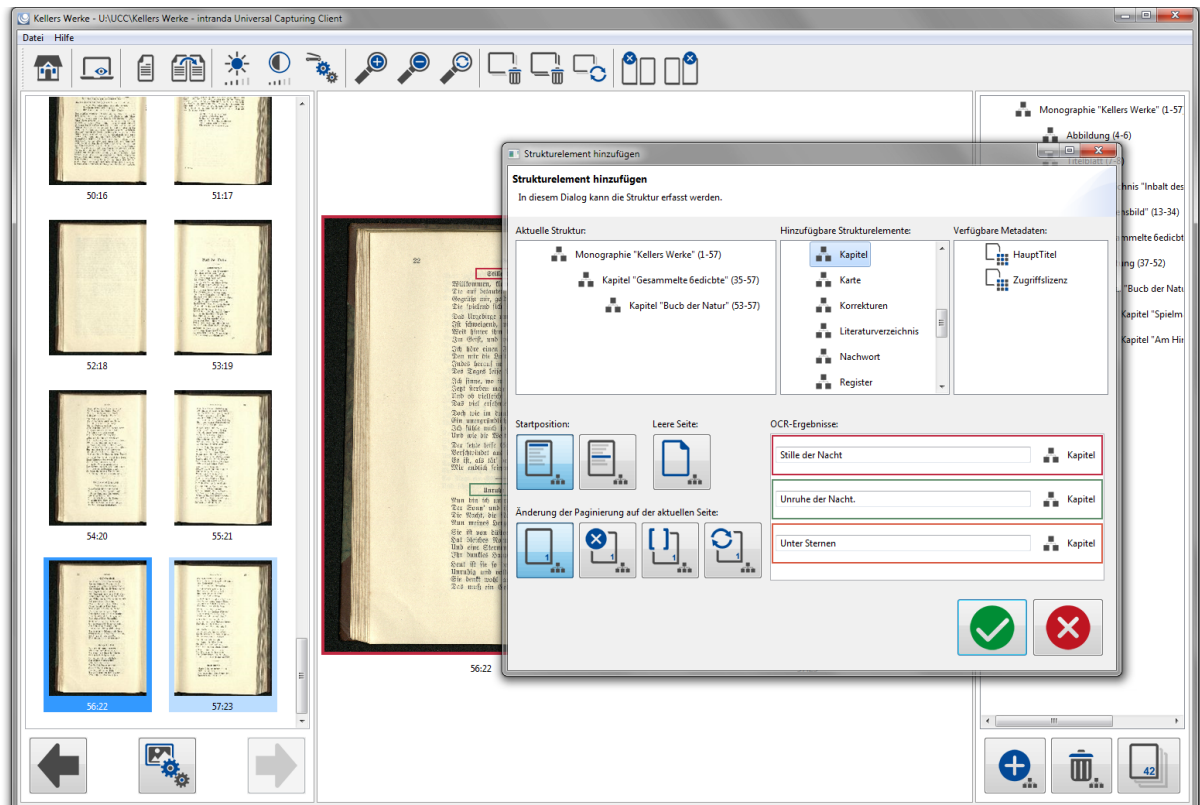
Figure 27: Generating multiple structure elements from a bordered item in the scanned image

The UCC allows you to subsequently delete structure elements if required. To do this, select the structure element from the right hand side and click the `Structure element` icon. This will open a dialogue in which you can delete the structure element in question.
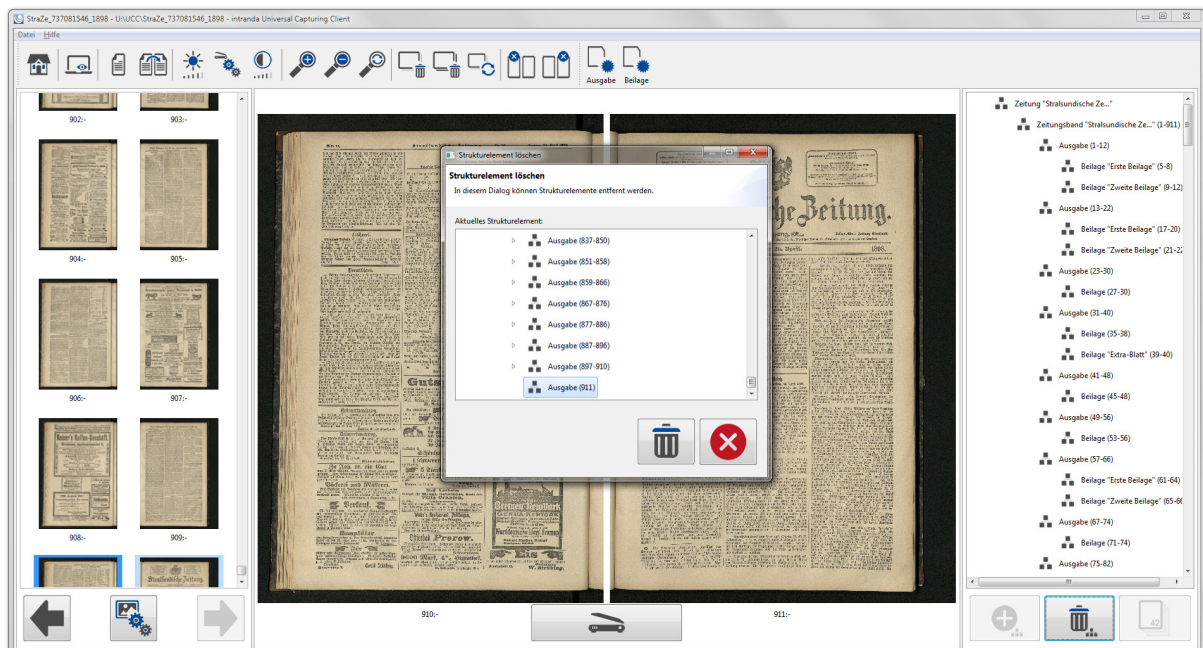

Figure 28: Deleting a structure element

4.3.4 Pagination

As well as scanning the source work, you can assign a pagination. To open the UCC's dedicated pagination dialogue, select the page for which you want to assign a pagination by clicking on that page. Next click the pagination icon to open the pagination dialogue.
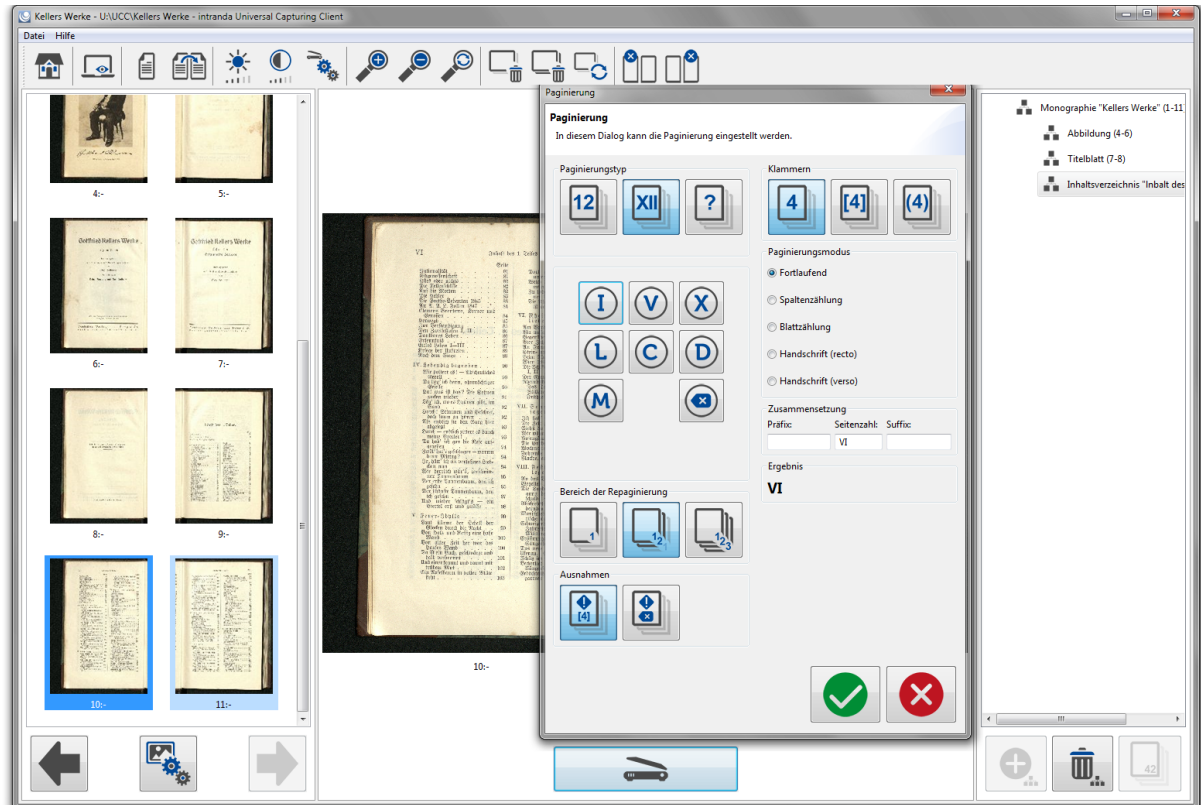

Figure 29: Specifying the pagination

At this point you can specify the pagination details. Select the pagination type in the top left of the window to set Arabic, Roman or no pagination. Further detailed pagination settings are available on the right-hand side of the dialogue box, e.g. use of brackets or fictitious pagination. You can also assign individual prefixes and suffixes. Before applying the settings, you can check the result of your options on the right-hand side of the dialogue box.

In the bottom-left corner of the dialogue box you can tell the UCC what to do with subsequent pages. For example, you can instruct the program to assign a new pagination for each subsequent page, only to paginate the current page or to apply the new pagination up to the next change to avoid subsequently modifying the book's existing pagination.

4.4 Quality control (correcting OCR results)

When you have finished scanning the source material and structured the resulting digitised images, you can move on to the next active workflow step on the UCC start page. This involves quality control and correcting the OCR results generated from the bordered items in the images. To run OCR correction, select the `Quality control` icon on the start page.

Figure 30: Selecting the quality control icon on the start page

When you select this icon, the UCC will open quality control view, which is divided into three sections. The section on the left contains all the image fragments generated when you placed borders around certain items. Next to each fragment is the resulting OCR text. If you click inside one of the boxes containing OCR text, the entire image will be shown on the right so that you can see the OCR results for that image fragment in context when editing.
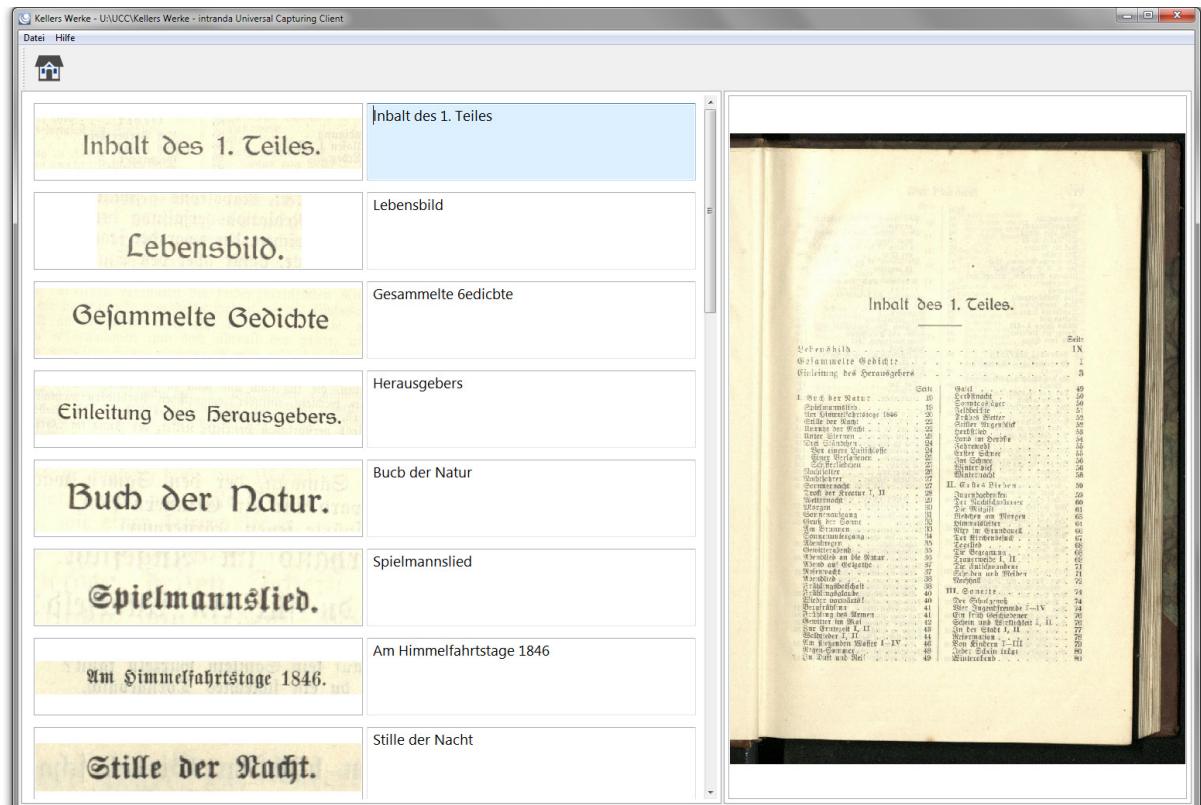
Figure 31: Editing window for correcting OCR results

You can use the tab key on your keyboard to move between the various OCR text boxes and edit the results as required. If you have generated a very large number of image fragments, you can use the buttons at the bottom of the window to display the next set of image fragments.

This makes the process of correcting OCR results very fast and straightforward. Once you have made the necessary corrections, the process of scanning the source work and capturing data is complete, and you can export or publish your results.

4.5 Publishing your results

When you have completed all the individual steps of the scanning project, you can go on to publish your data. To do this, simply click the `Publish` icon on the start page.

Figure 32: Selecting the publish icon on the start page

This will open a small dialogue box in which you can specify the location to which you want to export your results. A number of options are available.
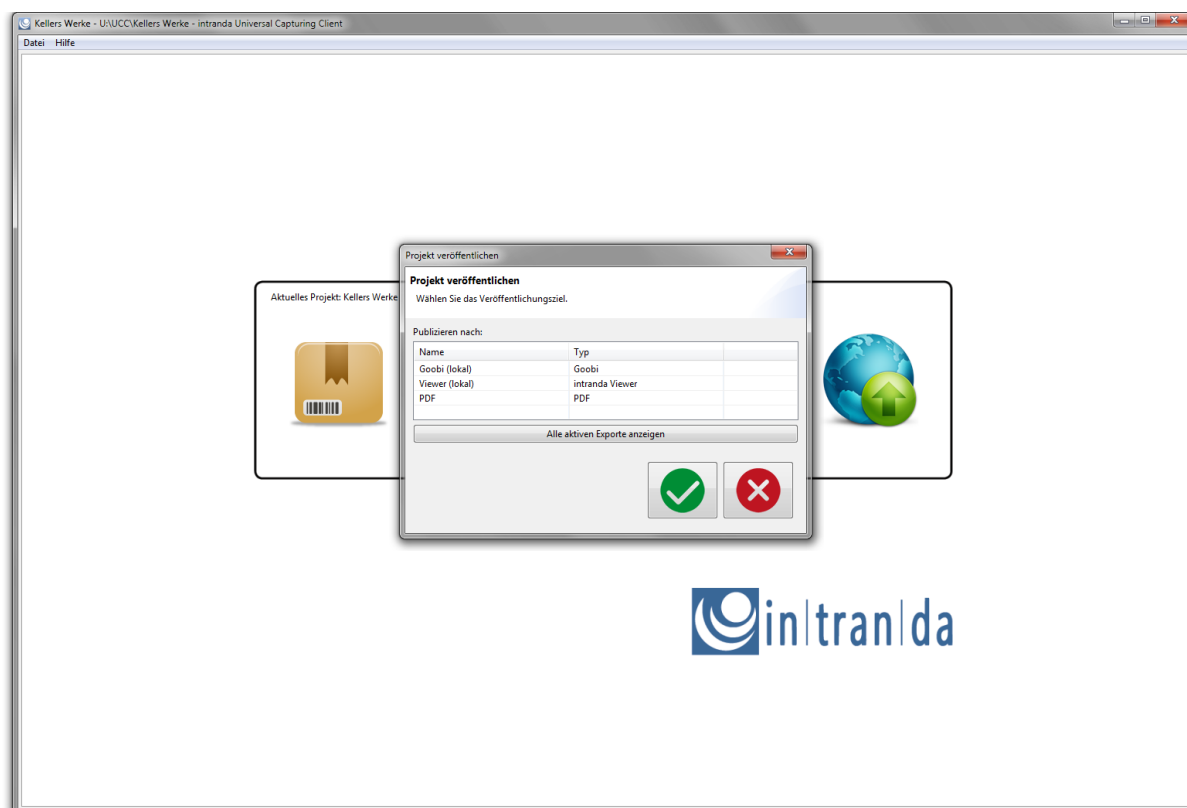
Figure 33: Dialogue box for publication of the digitised material

Depending on the way you have set up your project, you can select a number of export destinations. To export the results in PDF format, simply choose the PDF option and select a target directory.

If you want to publish your results in the intranda viewer, the UCC will export both your digital images and a METS file to the intranda viewer, which will then start indexing the data and adding the digitised material to its repository.

If you instruct the UCC to export your results to Goobi, the data will be exchanged with Goobi in two steps. First the METS file will be sent to Goobi via the Web API. Next, the digital images will be copied to the work directory designated by Goobi. This is usually integrated as a network drive and must be available when exporting to ensure that the data are transferred correctly.

5 Contact

Intranda partner	Contact
Florian Alpers	intranda GmbH
Steffen Hankiewicz	Bertha-von-Suttner Str. 9
Vanessa López Campo	D – 37085 Göttingen
	http://www.intranda.com
	support@intranda.com