



DpuScan

Janich & Klass
Computertechnik GmbH



DpuScan 6.x

Referenzhandbuch

Copyrights

© 1997 bis 2021 Janich & Klass Computertechnik GmbH. Alle Rechte vorbehalten. Gedruckt in Deutschland. Die in dieser Dokumentation enthaltenen Informationen sind Eigentum der Janich & Klass Computertechnik GmbH. Ohne schriftliche Genehmigung der Janich & Klass Computertechnik GmbH begründen weder der Empfang noch der Besitz dieser Informationen irgendein Recht auf Reproduktion oder Veröffentlichung irgendwelcher Teile davon.

Warenzeichen

Alle Produktnamen und Logos sind Warenzeichen oder eingetragene Warenzeichen der jeweiligen Eigentümer.

Haftungsausschluss

Die Anweisungen und Beschreibungen in diesem Handbuch waren zum Druckzeitpunkt zutreffend. Wir behalten uns jedoch das Recht vor, sowohl Beschreibung als auch Produkt jederzeit ohne Benachrichtigung zu ändern. Nach dem derzeitigen Stand der Softwaretechnik ist es nicht möglich, Programme zu entwickeln, die unter allen Bedingungen in jeder Konfiguration fehlerfrei arbeiten. Die Janich & Klass Computertechnik GmbH übernimmt keinerlei Haftung für Defekte, die direkt oder indirekt durch Fehler dieses Handbuches, Weglassen von Informationen oder durch Unstimmigkeiten zwischen diesem Referenzhandbuch und dem Produkt entstanden sind.

Aktualität

Es ist möglich, dass im Internet eine neuere Version dieses Handbuches verfügbar ist. Wir empfehlen deshalb, die Version anhand des auf dieser Seite abgedruckten Datums mit der Version auf dem Internet zu vergleichen. Falls die Version im Internet neueren Datums ist, sollten Sie diese herunterladen und ggf. selbst ausdrucken.

Die aktuelle Version des DpuScan Referenzhandbuch finden Sie im Web unter:

<http://www.jkimaging.com/pdf/DpuScan-Referenzhandbuch.pdf>

Inhaltsverzeichnis

1 OCR FineReader 6.0 Plugin	4
1.1 General Functions	4
1.2 OCR Area Search	4
1.3 OCR Fulltext Export	5

1 OCR FineReader 6.0 Plugin

With the OCR FineReader 6.0 Plugin, DpuScan can perform an OCR Search of search areas or fulltext export. Performing area search, a section of the image is handled and the recognized text is passed to a variable, handled in DpuScan. In fulltext search the whole image is handled and the recognized formatted text is exported to a file. Export formats for example are rtf, pdf or html. The plugin based on the FineReader 6.0 Engine. The following links describe the settings of the plugin, adjusted in the setup dialog.

General Functions

OCR Area Search

OCR Fulltext Export

1.1 General Functions

Taskmode	The plugin is called by DpuScan tasks. The Ocr Search mode (area search or fulltext search) can be defined with this combo box.
Load file	Loads a file from harddisk. All established file formats are supported.
Add area	Creates a frame in the loaded image. The size and the position of the frame can be changed subsequently.
Delete Area	Deletes the active frame.
Test Area	Performs an Ocr area search with the active document in the Dpu Window. All defined areas are analysed. The results are displayed in a dialog box.
Test Fulltext	Performs an Ocr area search with the active document in the Dpu Window. The result is exported in to a file in the current export format. A dialog box offers a direct control of the created file.
OK	Closes the dialog box and stores all settings.
Cancel	Closes the dialog box without saving.
Help	Opens this help screen.

1.2 OCR Area Search

Defined Areas	The defined areas are numbered in the listbox. Clicking one of the elements of the listbox marks the appropriate frame in the preview window. The position and the size of the frame and the name of the appropriate variable are displayed in edit fields. The Ocr Area result is stored in this variable.
Left	Depending on the alignment of the frame (Position relative to) the position of the left margin of the frame is displayed in this edit field. The measurement of the value is tenth of a millimeter.
Top	Depending on the alignment of the frame (Position relative to) the position of the top margin of the frame is displayed in this edit field. The measurement of the value is tenth of a millimeter.

Width	Defines the width of the frame. The measurement of the value is tenth of a millimeter.
Height	Defines the height of the frame. The measurement of the value is tenth of a millimeter.
Position relative to	The position of a frame attributes to one of the four corners of a document. This is important to find text in documents with different sizes (for example, if the text is always in the bottom/right corner).
Percent Code	Using the search on parts of the documents by defining frames, the results of the Ocr recognition are stored in variables. These variables are used in many dialogs of DpuScan. The names of these variables are defined in this edit field.
Linefeeds	Linefeeds and paragraph separation in the OCR area search can be replaced by spaces or CTRL-A characters (0x1) or they stay as linefeeds (0xa) in the text. This option defines the which method is taken. Alternatively linefeed can be deleted.
Paragraph separation	This option enables or disables the paragraph separation. Paragraphs are not the same as linefeeds, which may be in paragraphs additionally. With enabled option the paragraph separation are performed in the way, that is defined in Linefeeds.
Language	Choose between different languages for the OCR process. Languages could have different characters, so the results of the OCR process may be different by choosing different languages.
Text Type	The Ocr search distinguishes between different text types. The "normal" parameter is the default. Additionally there is typewriter or matrix. The options OCR-A, OCR-B und MICR-E-13B are special fonts, used on checks. They have certain letters and a specific defined appearance.

1.3 OCR Fulltext Export

Language	Choose between different languages for the OCR process. Languages could have different characters, so the results of the OCR process may be different by choosing different languages.
Text Type	The Ocr search distinguishes between different text types. The "normal" parameter is the default. Additionally there is typewriter or matrix. The options OCR-A, OCR-B und MICR-E-13B are special fonts, used on checks. They have certain letters and a specific defined appearance.
Export Format	The file format for a OCR fulltext search can be defined with this parameter. The result of the OCR search is stored in a file with the specified extension in the directory of the current image. Choose one of the following formats: RTF, HTML, XLS, PDF, DBF and TXT
Save Mode	Specifies whether only text or additionally image elements should be saved in the exported file.

Only Text: Only text elements are stored (default setting).

Text and pictures: Text elements are stored as text and image elements are stored as images.

	Pictures over Text: (Only in Pdf Format): The text is stored hidden in the background. The images are placed in the foreground. This output is similar to the original.
Formfeed in textformat	If several images are stored in one file (multistreams), the associated ocr fulltext results are stored in one file, too. In text format you can define a formfeed to separate the sheets. In all other formats this is the default behavior.
Keep lines in RTF format	Specifies whether line breaks should be saved in the exported file.
Correct skew	Activate this checkbox for deskewing image during OCR process.
Invert image	Activate this checkbox for inverting image during OCR process.
Despeckle	Activate this checkbox for cleaning image during OCR process.
Correct Orientation	Recognizes the text orientation of the source image. The resulting document, created by the fulltext search, is rotated to the correct orientation, if this checkbox is activated. The recognized text is displayed correctly.