



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 Allgemeine Einstellungen	4
1.2 OCR Feldsuche	4
1.3 OCR Volltext Export	5

1 OCR FineReader 6.0 Plugin

Mit dem OCR FineReader 6.0 Plugin kann in DpuScan eine OCR Suche von Teilbereichen oder ganzen Textseiten durchgeführt werden. Bei der Teilbereichssuche oder Feldsuche wird ein Ausschnitt von einer Bild durchsucht und die dort gefundenen Texte in einer DpuScan Variable zurückgegeben. Die Volltextsuche untersucht alle Bereiche eines Bildes und kann die Ergebnisse formatiert ausgeben. Als Exportformat stehen rtf, html, pdf, xml und andere Formate zur Verfügung. Das Plugin basiert auf der FineReader 6.0 Engine. Die folgenden Links verweisen auf die Einstellmöglichkeiten des Plugins, die Sie im Einstelldialog durchführen können.

Allgemeine Funktionen
OCR Feldsuche
OCR Volltext Export

1.1 Allgemeine Einstellungen

Taskmodus	Das Plugin wird in der Task von DpuScan aufgerufen. Die Art der Ocr Suche (Feldsuche oder Volltextsuche) kann über dieses Kombinationsfeld festgelegt werden.
Datei laden	Lädt eine Datei von der Festplatte. Es werden alle gängigen Dateiformate unterstützt.
Feld hinzufügen	Erzeugt einen neuen Rahmen. Sie können den Rahmen in seiner Größe verändern oder an eine andere Position schieben. Halten Sie die linke Maustaste gedrückt während Sie sich im Rahmen befinden, um den Rahmen zu verschieben. Wenn Sie auf den Rand des Rahmens klicken, können Sie die Größe des Rahmens verändern.
Feld löschen	Löscht den in der Box Definierte Felder markierten Rahmen beziehungsweise das blaue Rechteck aus dem Fenster.
Test Feld	Führt eine OCR Feldsuche mit dem gerade im Fenster dargestellten Bild durch. Dabei wird in allen definierten Bereichen gesucht. Die Ergebnisse werden in einer Dialogbox dargestellt.
Test Voll	Führt eine OCR Volltextsuche mit dem gerade im Fenster dargestellten Bild durch. Der Export des Ergebnisses erfolgt in dem gerade eingestellten Exportformat. Eine Dialogbox ermöglicht die direkte Kontrolle des erzeugten Formats.
OK	Schließt die Dialogbox und speichert alle Definitionen auf Festplatte.
Abbrechen	Schließt die Dialogbox ohne zu speichern.
Hilfe	Öffnet diesen Hilfebildschirm.

1.2 OCR Feldsuche

Definierte Felder	Die definierten Felder werden in dem Listefeld durchnummeriert. Ein Klicken auf eines der Elemente des Listefeldes markiert den zugehörigen Rahmen im Vorschaufenster und zeigt die Position und Größe in den vier Editfeldern an sowie den Namen der zugehörigen Prozentcode Variable. Das Ergebnis der Feldsuche wird in dieser Prozentcode Variable gespeichert.
Links	Gibt abhängig von der Ausrichtung des Rahmens (Position relativ zu) die Position des linken Rands des Rahmens an. Die Maßeinheit des Wertes ist Zehntelmillimeter.

Oben	Gibt abhängig von der Ausrichtung des Rahmens (Position relativ zu) die Position des oberen Rands des Rahmens an. Die Maßeinheit des Wertes ist Zehntelmillimeter.
Breite	Gibt die Breite des Rahmens an. Die Maßeinheit des Wertes ist Zehntelmillimeter.
Höhe	Gibt die Höhe des Rahmens an. Die Maßeinheit des Wertes ist Zehntelmillimeter.
Position relativ zu	Die Position der linken oberen Ecke der einzelnen Felder kann relativ zu einer der vier Ecken des Blattes gewählt werden. Dadurch können auch Texte auf Blättern unterschiedlicher Größe gefunden werden, wenn sie zum Beispiel immer unten rechts stehen.
Prozentcode	Definiert den Namen des Prozentcodes, der an DpuScan übermittelt wird. Durch Zugriff auf die Variable kann in DpuScan der Wert der Feldsuche beispielsweise im Infofenster oder in der Protokolldatei ausgegeben werden.
Zeilenumbrüche	Zeilenumbrüche und Absatzumbrüche in der OCR Feldsuche können durch Leerzeichen oder CTRL-A Zeichen (0x1) ersetzt werden oder als Zeilenvorschub (0xa) belassen werden. Diese Option legt fest welche der oben aufgeführten Methoden durchgeführt wird. Alternativ dazu können Zeilenumbrüche gelöscht werden.
Absatztrennung	Diese Option schaltet die Absatztrennung ein oder aus. Absätze sind keine Zeilenumbrüche die es zusätzlich noch geben kann. Bei eingeschalteter Absatztrennung werden die Absatzumbrüche (wie auch die Zeilenumbrüche), so wie es in Zeilenumbrüche eingestellt wurde, ersetzt oder belassen.
Sprache	Es können verschiedene Sprachen ausgewählt werden, mit denen die OCR Suche durchgeführt wird. Sprachsysteme bauen auf verschiedenen Buchstaben oder Satzzeichen auf. Das Ergebnis einer OCR Suche hängt somit auch von der gewählten Sprache ab.
Texttyp	Sie können den Texttyp der zu suchenden Texte einstellen. Die Einstellung "normal" ist die Standardeinstellung. Außerdem können noch Schreibmaschine oder Matrix eingestellt werden. Die Optionen OCR-A, OCR-B und MICR-E-13B sind spezielle Schriftsymbole auf Schecks, die sowohl einen bestimmten Zeichensatz als auch ein genau definiertes Aussehen haben.

1.3 OCR Volltext Export

Sprache	Es können verschiedene Sprachen ausgewählt werden, mit denen die OCR Suche durchgeführt wird. Sprachsysteme bauen auf verschiedenen Buchstaben oder Satzzeichen auf. Das Ergebnis einer OCR Suche hängt somit auch von der gewählten Sprache ab.
Texttyp	Sie können den Texttyp der zu suchenden Texte einstellen. Die Einstellung "normal" ist die Standardeinstellung. Außerdem können noch Schreibmaschine oder Matrix eingestellt werden. Die Optionen OCR-A, OCR-B und MICR-E-13B sind spezielle Schriftsymbole auf Schecks, die sowohl einen bestimmten Zeichensatz als auch ein genau definiertes Aussehen haben.
Export Format	Für eine OCR Volltext Suche können verschiedene Dateiformate ausgewählt werden. Das Ergebnis der OCR Suche wird in einer Datei mit der entsprechenden Dateierweiterung im Verzeichnis des zu

	bearbeitenden Bildes gespeichert. Folgende Export Formate können gewählt werden:
	RTF, HTML, XLS, PDF, DBF und TXT
Speicher Modus	Legt fest, ob nur der Text oder auch Bildelemente in der Ausgabedatei mitgespeichert werden.
	Nur Text: Es werden ausschließlich Textelemente übernommen (Standardeinstellung)
	Text und Bilder: Textelemente werden als Text, Bildelemente als Bilder gespeichert
	Bilder über Text: (Nur im Pdf Format): Der Text wird unsichtbar im Hintergrund gespeichert, die Bildelemente legen sich darüber. Diese Ausgabe entspricht vom Erscheinungsbild genau dem Original.
Seitenumbruch im Textformat	Werden mehrere Bilddateien in einer Datei gespeichert (Multistreams), dann werden auch mehrere OCR Volltext Suchergebnisse in einer zugehörigen Textdatei gespeichert. Im Textformat haben Sie die Wahl die einzelnen Seiten durch einen Seitenumbruch zu trennen. In allen anderen Formaten geschieht das automatisch.
Zeilenumbrüche im RTF Format beibehalten	Legt fest, ob die Zeilenumbrüche mit in die Ausgabedatei übernommen werden.
Geraderücken	Aktivieren Sie diese Option, um das Bild während der Bearbeitung gerade zu rücken.
Bild invertieren	Aktiviert/deaktiviert den Invertiermodus während die Datei bearbeitet wird.
Bild säubern	Aktiviert/deaktiviert das Säubern von Verschmutzungen während der Bildbearbeitung.
Bildausrichtung korrigieren	Erkennt, ob eine Seite um 90 oder 180 Grad rotiert werden muss, um den Text korrekt zu erkennen. Die durch die Volltextsuche erzeugte Datei ist dann korrekt rotiert und die gefundenen Textelemente werden korrekt angezeigt.