

OCR mit Bluebeam Revu eXtreme

Die Optical Character Recognition – kurz OCR – oder auch Texterkennung genannt, erlaubt die Umwandlung von eingescannten PDF Dokumenten in durchsuchbare PDF.

Unter der Texterkennung versteht man eine Umwandlung von Bildinformationen in echten Text. Um die Bildinformationen von gescannten Dokumenten zur Weiterverarbeitung und Wiederverwendung in Texte umzuwandeln, steht Ihnen folgende Vorgehensweisen zur Verfügung.

Erstellung von durchsuchbare PDF-Dateien

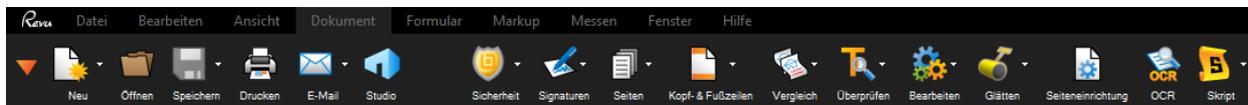
Eingescannte PDF-Dateien von Zeichnungen oder anderen Dokumenten sind zunächst einfache Kopien des Originaldokumentes. Diese enthalten ausschließlich Pixel und sind somit gerastert. Mit der OCR können diese Rasterinformationen – welche augenscheinlich Texte repräsentieren – in echten Text umgewandelt werden. Die so erzeugten Textinformationen werden versteckt hinter der eigentlichen Information gelegt und können nun über die Textsuche gefunden werden. Der Benutzer sieht in dieser Datei lediglich das, was auch eingescannt wurde. Die Erkennung von Schriften, Layouts sowie Layout-Elemente sind für die Texterkennung belanglos und werden ignoriert.

Die umgewandelten PDF-Dateien können auch mit dem Windows Explorer durchsucht werden; auf Webseiten gelegt sind die Seiten sogar von Suchmaschinen, welche die Dokumente indiziert haben auffindbar. Es ist auch möglich diese in diversen Archiv und EDM-Systemen zu verwenden, auch hier kann man in den Dokumenten suchen.

Wichtig: OCR ist nur in Bluebeam Revu eXtreme verfügbar! Menü und Werkzeugleisteneinträge sind nicht in den Versionen Standard und CAD verfügbar

OCR-über ein Dokument laufen lassen

1. Öffnen Sie das Dokument über das die OCR gehen soll



2. Klicken Sie auf **Dokumente** und anschließend auf **OCR**; alternativ können Sie auch die Tastenkombination Strg + Shift + O verwenden, es öffnet sich eine Dialogbox
3. Wählen Sie unter Erkennungssprache die verwendete Sprache aus, es können mehrere Spracherkennungen zeitgleich verwendet werden
4. Setzen Sie die OCR-Konfigurationen wie gewünscht



- *Verzerrung korrigieren:* Korrektur von Winkelabweichungen im Dokument
- *Ausrichtung erkennen:* Automatische Erkennung der Seitenausrichtung
- *Text in Bildern erkennen:* Texte werden in Bildern erkannt
- *Dokumententyp:* Voreinstellung um welche Art von Dokument es sich handelt; die Einstellung CAD-Zeichnungen tendiert dahin Textformatierungen zu ignorieren
- *Optimiert für:* Auswahl zwischen Schnell und ungenauer oder Genau und dafür langsamer
- *Vertikalen Text erkennen:* Vertikal erkannter Text wird automatisch gedreht
- *Größe des Seitenabschnitts:* Maximale Seitenanzahl die zeitgleich vom OCR bearbeitet werden kann
- *Vektorseiten überspringen:* Vektorbasierte Seiten werden übersprungen
- *Max. Vektorgröße:* Maximale Vektorgröße die während des OCR-Prozess analysiert wird; Vektoren die größer sind als eingestellt werden in der Vorbearbeitung verworfen. Herabsetzen dieses Wertes kann die Geschwindigkeit vergrößern, kann aber auch mehr Text zur Folge haben

5. Legen Sie den Seitenbereich fest (Alle Seiten, Aktuell, Benutzerdefinierter Bereich)
6. Klicken Sie auf OK; OCR wird nun ausgeführt

OCR über mehrere Dokumente laufen lassen

1. Arbeiten Sie die Punkte 1–5 bei **OCR über ein Dokument laufen lassen** ab
2. Klicken Sie auf Dateien hinzufügen, eine neue Dialogbox öffnet sich
3. Wählen Sie Ihre Einstellungen wie Gewünscht aus; die Einstellungen können für jede PDF separat vorgenommen werden
 - Dateiliste: Über Geöffnete Dateien hinzufügen können Sie bereits offene Dokumente zur Verarbeitung ergänzen; über Hinzufügen haben Sie die Möglichkeit Dokumente über den Windows Browser hinzuzufügen
 - Seitenbereich: für jedes Dokument kann ein Seitenbereich ausgewählt werden (Alle Seiten und Benutzerdefinierte Bereiche)
 - Seitenfilter: Legt fest ob auf Geraden und/oder ungerade Seiten OCR angewendet wird sowie Seiten im Hoch- und/oder Querformat
4. Klicken Sie auf OK und sie gelangen zur vorherigen Dialogbox zurück; klicken Sie hier auf OK und der Bearbeitungsvorgang wird gestartet

