

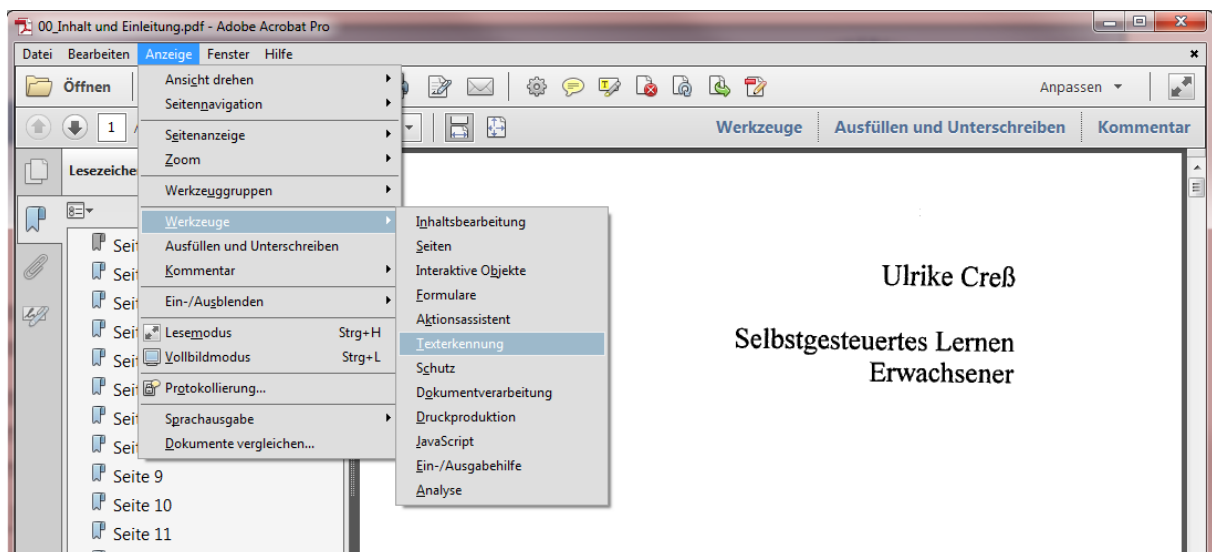
Texterkennung für gescannte PDF-Dokumente

PDF-Dateien, die durch das Abscannen von gedruckten Vorlagen erzeugt wurden, sind zunächst bildgenaue Reproduktionen des Originaldokuments. Sie enthalten jedoch keine Text- sondern ausschließlich „Pixelinformationen“. Durch optische Zeichenerkennung oder OCR (Optical Character Recognition) können die Pixelinformationen in Text umgewandelt werden. Die Textinformationen werden dann in den PDF-Dateien, quasi in einer „zweiten Ebene“ hinter die sichtbaren Bildinformationen gelegt. Mit Hilfe der Textsuche im Adobe Reader sind die Dokumente dann durchsuchbar und es können Stellen markiert und kopiert werden, um sie in eigene Dokumente einzufügen.

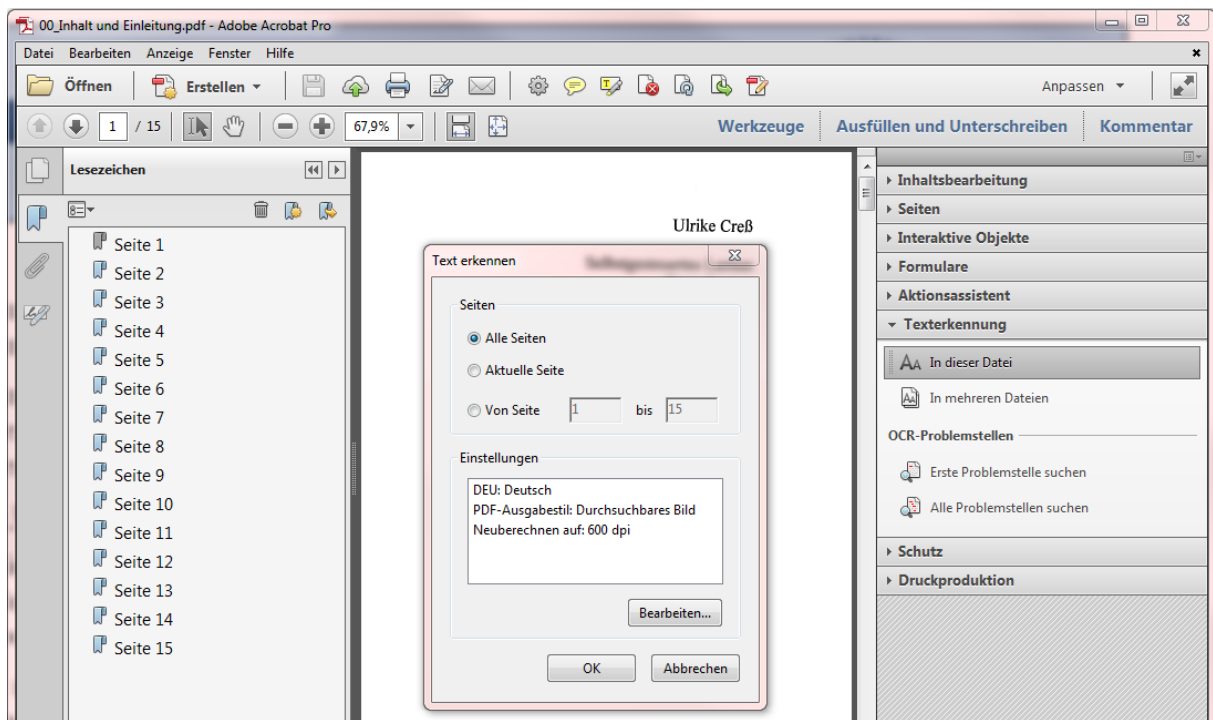
Empfohlene Vorgehensweise

Studenten und Mitarbeitern steht an den Computern der Universität Bamberg das Programm **Adobe Acrobat XI Pro** zur Verfügung. Dieses Programm bietet neben umfangreichen Möglichkeiten zur Erstellung und Bearbeitung von PDF-Dateien auch die Möglichkeit zur Texterkennung:

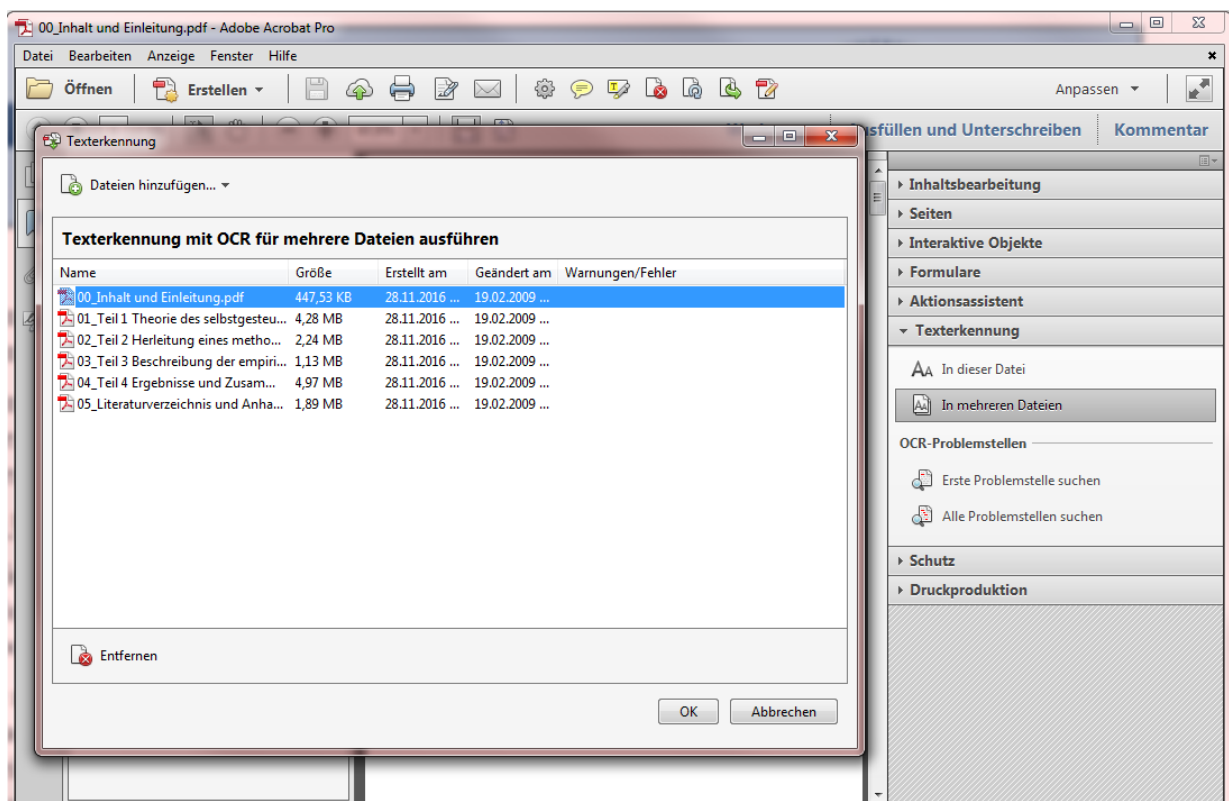
1. Öffnen Sie die PDF-Datei in Adobe Acrobat Pro.
2. Wählen Sie im Menü „Anzeige“ den Befehl „Werkzeuge“ > „Texterkennung“. Das Bedienfeld „Texterkennung“ wird geöffnet.



3.1. Klicken Sie auf „In dieser Datei“, falls Sie nur in der aktuellen Datei OCR durchführen lassen wollen. Geben Sie die entsprechenden Seiten an. Um das Ergebnis der Texterkennung zu verbessern ist es wichtig, dass Sie gegebenenfalls die Sprache auf die Sprache der Vorlage umstellen. Klicken Sie auf „OK“. Die optische Zeichenerkennung wird jetzt auf das gescannte Dokument angewendet.



3.2 Wenn Sie mehrere Dokumente auf einmal bearbeiten wollen wählen Sie alternativ „In mehreren Dateien“ aus. Fügen Sie alle zu bearbeitenden Dateien hinzu und wählen Sie dann „OK“.



Legen Sie in den Ausgabeoptionen fest, wo die bearbeiteten Dateien nach der Fertigstellung abgespeichert werden, welchen Namen die Dateien bekommen und ob die vorhandenen Dateien überschrieben werden sollen. Klicken Sie „OK“.

Wählen Sie bei den Texterkennungseinstellungen die Sprache der Vorlage aus. Die restlichen Einstellungen müssen nicht angepasst werden. Nach Klick auf „OK“ startet die optische Zeichenerkennung. Je nach Umfang des Auftrags kann dieser einige Zeit in Anspruch nehmen.