

OCR

Texte mit Epson-Scan einscannen und in Word weiterverarbeiten

von Bernhard Koch
im XP-Forum

Vorgang: Der Text wird eingescannt, zuerst in eine tif-Datei umgewandelt und dann von der OCR-Funktion von MS Office Document Imaging in ein Word-Dokument transformiert. Die Formatierung bleibt dabei erhalten.

Bedienschritte:

- 1.) Mit dem Programm Epson Scan das Dokument einscannen.
Einstellungen:
 - Standard Modus
 - Text/Strichzeichnung
 - Schwarzweiß
 - Benutzerdefiniert
 - Auflösung mind. 300 dpi, besser 600 dpi je nach Güte der Vorlage
- 2.) Einstellungen bei „Einstellung zum Speichern der Dateien“:
 - Ort: Speicherort angeben, sonst wird die Datei automatisch unter C:/Dokumente und Einstellungen, „Benutzername“, Eigene Dateien, Eigene Bilder abgelegt.
 - Dateiname: Eigenen Dateinamen vergeben sonst wird automatisch einer vergeben und man muss die Datei suchen.
 - Bildformat: Auswählen Multi-TIFF (*.tif).
- 3.) Nach dem Scanvorgang im Fenster „Epson Scan“ auswählen, ob noch weitere Seiten gescannt werden sollen. Wenn nicht, dann bei „Seite wird bearbeitet“ mit „ok“ abschließen.
- 4.) Anschließend erscheint ein Fenster mit einer Dateiübersicht. Hier die gewünschte tif-Datei öffnen.
- 5.) Das geöffnete Dokument mit der rechten Maustaste anklicken um das Kontextmenu aufzurufen. Dort „Öffnen mit“ anklicken. Im neuen Menu „Microsoft Office Document Imaging“ auswählen. Es erscheint das Fenster „MS Office Document Imaging“
- 6.) In der Befehlszeile oben das vorletzte Symbol „Texterkennung mittels OCR“ anklicken. Dann das letzte Symbol „Text an Word senden“ anklicken und mit „ok“ bestätigen.
- 7.) Im Fenster „Microsoft Word“ erscheint jetzt das umgewandelte Dokument. Es sollte erst abgespeichert und dann bearbeitet werden.
Beim Abspeichern über „Datei“, „Speichern unter“ nicht vergessen bei „Dateityp“ „Word Document (.doc)“ auszuwählen.
- 8.) Zur anschließenden Bearbeitung aktiviert man die automatische Fehleranzeige über „Extras“, „Optionen“, „Rechtschreibung und Grammatik“; mit „Dokument überprüfen“ bestätigen.
Wenn bestimmte Buchstaben, Silben oder Worte falsch wiedergegeben sind, kann man über „Bearbeiten“, „Suchen“ und „Ersetzen“ den Text schnell bereinigen. Funktioniert aber nicht immer, evtl. hilft das Umkopieren in einen neues Word Dokument.

Informationen zu Microsoft Office Document Imaging

In Microsoft Office Document Imaging können Sie ein gescanntes Dokument genauso problemlos verwenden wie andere Microsoft Office-Dokumente auf dem Computer.

So können Sie beispielsweise:

- Einseitige und mehrseitige Dokumente scannen.
- [Optische Zeichenerkennung \(Optical Character Recognition oder OCR\)](#) ausführen.
- Gescannte Texte und Bilder in Microsoft Word kopieren und exportieren.
- Ein gescanntes Dokument genauso problemlos lesen wie ein Papierdokument.
- Text in gescannten Dokumenten suchen.
- Gescannte Dokumentseiten genauso problemlos umstrukturieren wie Dokumente in einem Ordner erneut anordnen.
- Gescannte Dokumente per E-Mail oder als Fax über das Internet an andere Benutzer senden.
- Anmerkungen zu gescannten Dokumenten und Faxen hinzufügen.

Wann sollten Sie Microsoft Office Document Imaging verwenden?

Verwenden Sie Microsoft Office Document Imaging zum Ausführen folgender Aktionen:

- Zugreifen auf und Arbeiten mit dem Text oder den Bildern, die in gescannten oder Faxdokumenten enthalten sind.
- Extrahieren von Text aus vorhandenen gescannten Dokumenten oder Faxdokumenten zwecks Archivierung.
- Erneutes Anordnen der Reihenfolge von Seiten in mehrseitigen gescannten Dokumenten.
- Schnelles und problemloses Lesen gescannter Dokumente auf dem Bildschirm.
- Hinzufügen von Anmerkungen zu einem elektronischen Fax und Zurücksenden des Fax an den Absender.

Scanning und Imaging

Microsoft Office Document Imaging umfasst zwei Komponenten (eine Komponente zum Scannen und eine für die Bildbearbeitung), die im Startmenü von Windows separat als **Microsoft Office Document Scanning** und **Microsoft Office Document Imaging** aufgeführt sind.

Mit der ersten Komponente wird das Scannen von Dokumenten auf dem Computer unter Verwendung eines installierten Scanners gesteuert. Dies erfolgt durch das Verwenden von Scannvoreinstellungen, mit denen der Scanner mithilfe von Einstellungen gesteuert wird, die für bestimmte Zwecke optimiert wurden. Die Scannvoreinstellung **Schwarzweiß** ist beispielsweise optimal für das Scannen von Textseiten für die OCR geeignet, während die Scannvoreinstellung **Farbe** am besten für das Scannen von Farbfotos oder mehrfarbigem Bildmaterial geeignet ist. Darüber hinaus wird die OCR standardmäßig sofort nach dem Scannvorgang automatisch in Textdokumenten ausgeführt, und Sie können mehrere Seiten problemlos zusammen in eine einzige Datei scannen.

Mit der zweiten Komponente können Sie problemlos gescannte Dokumente auf dem Bildschirm anzeigen, mehrseitige Dokumente neu anordnen, erkannten Text markieren und bearbeiten, Anmerkungen zu gescannten Dokumenten und elektronischen Faxen hinzufügen sowie Dokumente per E-Mail oder Fax an andere Benutzer senden.

Ratschläge aus Benutzerforen:

Mit "Microsoft Office Document Imaging" kannst du ein Papierdokument einscannen und dabei gleich für die Texterkennung aufbereiten.

Bereits eingescannte Abbilder eines Dokuments kannst du auch nachträglich mit dem OCR-Modul bearbeiten. Dazu gibt es den Befehl « Extras - Texterkennung mit OCR » (Version 2003) bzw. « Datei-Texterkennung mittels OCR » (Version XP). Nach der Texterkennung lassen sich die Textelemente mit der [Maus](#) markieren und über die Zwischenablage als unformatierter Text in ein Word-Dokument kopieren. Oder du nutzt den Befehl « Text an Word senden »

OCR für einen vorhandenen Scanner

Wer einen Scanner hat, aber keine OCR (Optical Character Recognition = Texterkennung) dazu, kann auf folgende Weise eine OCR von Office XP bzw. Office 2003 verwenden:

Die Office-Versionen XP und 2003 verfügen über eine eigene OCR-Software, mit der man eingescannte ***.tif-Dateien** lesbar machen kann. Damit kann man sich eventuell eine separate OCR-Software sparen. So gehen Sie vor:

1. "Start - MS Office - Office Tools - Office Document Scanning" anklicken.
 2. Falls dieser Programmbestandteil bei der Installation nicht mitinstalliert wurde, müssen Sie bei der entsprechenden Aufforderung die Office-CD einlegen und den Programmteil nachinstallieren.
 3. Wenn das passiert ist, klicken Sie den Menüpunkt "Neues Dokument scannen" an.
 4. Falls es noch nicht aktiviert ist, müssen Sie "schwarz-weiß" auswählen.
 5. Jetzt können Sie "Scannen" anklicken.
 6. Nach dem Scan-Vorgang erscheint das Fenster "MS Document Imaging".
 7. Jetzt importieren Sie den Text in MS Word durch Markieren in Imaging und Einfügen in ein leeres Word-Dokument.
 8. Speichern Sie dieses unter einem Namen ab.
 9. Nun können Sie alle Fenster wieder schließen; die Abfrage: "speichern in MS Office-Unterstützung TIF" kann mit nein beantwortet werden, da der Text ja in Word gespeichert ist.
- Hinweis: Bei Spaltentexten wird die Sache komplizierter, denn ein solcher kann nur spaltenweise nacheinander in eine Word-Datei importiert werden.

Beim ersten Scannen dauert der Vorgang recht lang, weil Office die Eigenschaften des Scanners prüft und abspeichert. Ab dem zweiten Scannen geht es wesentlich schneller.