

Scannereinsatz im Hausgebrauch

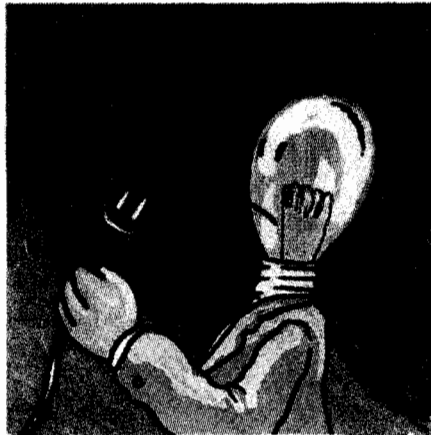
Einige Hinweise zur Auflösung und Weiterbearbeitung von Dokumenten

Schnell einmal über etwas drüberschauen, das bedeutet „scannen“ wörtlich. Tut man's am PC, ist das eine eher umständliche Sache. Trotzdem werden immer mehr Dokumente in den Computer eingescannt, schon um es auf der Festplatte „abgeheftet“ zu haben. Wir wollen hier einige Hinweise für den Hausgebrauch geben, nicht für den professionellen Einsatz von Scannern, etwa im Bereich der Fotografie. Scanner lassen sich auf unterschiedliche Abtastungen einstellen. Bevor sie gekauft werden, prahlen sie mit ihrer feinsten Feinheit, meist 1200 dpi, Dots per Inch. In der Praxis braucht man so feine Abtastungen kaum, ja, sie gelingen nicht einmal immer richtig. 1200 Punkte je Zoll sind 2,54 Mal weniger je Zentimeter: rund 500 je Zentimeter gleich 50 je Millimeter. Das ist fünfmal feiner als ein Haar. Wir haben deshalb zwar nicht immer mit der Standardauflösung von 100 dpi, aber doch mit 300 dpi beste Erfahrungen gemacht. Dazu kommt, dass bei höheren Auflösungen die Fokussierung des Scanners für Laien und Amateure oft nicht genau genug ist, um auch alle Feinheiten scharf abzubilden: Man erkennt auf einer mit 1200 dpi abgetasteten Briefmarke zwar das Druckraster, die unter einer Lupe gestochen scharfe Schrift bleibt im Scan aber trotzdem verwaschen. Bei Zeitungsartikeln reichen 200 dpi für anschließendes elektronisches Lesen (OCR-Erkennung, Optical Character Recognition). Vor allem sollte auf Schwarzweiß gestellt werden, statt bytespielig nicht vorhandene Farben einzulesen. Unsere PDF-Software richtet bei späterer optischer Buchstabenerkennung schräge Blätter gerade; sonst ist elektronisches Geraderücken von Eingescanntem eher mühsam, jedenfalls kniffliger, als die Vorlage gleich gut ausgerichtet schön gerade auf die Scannerplatte zu legen.

Alte Bücher sollte man ihrer Bindung halber nicht mit Gewalt auf die Glasplatte quetschen, Bibliotheken verbieten derartiges Bücherkopieren ganz. Die Bücher leiden, und bei allzu festem Druck schrammt gelegentlich sogar die Abtastleiste unter dem Glas. Buchseiten

„scannt“ man mit seinem digitalen Fotoapparat, den Band nur halb aufgeklappt, wobei man wie bei Gegenlichtaufnahmen die Belichtung erhöht oder einen extra Weißabgleich macht. Übrigens empfiehlt es sich auch bei älteren Scannern, mit einer höheren Belichtung zu experimentieren. Eine Buchseite im Format A5 mit 300 dpi gescannt gibt rund 4,3 Millionen Pixel; beim Fotoapparat-„Scannen“ sollte man demnach den Apparat auf wenigstens 5 Megapixel stellen. Dabei fotografieren man nicht weitwinklig, sonst zeigt die Seite kissenartige Verzerrungen.

Kamera-Zusatzsoftware kann nach dem Scannen gute Dienste tun. Hat eine



Vorlage Übergröße – etwa eine ganze Zeitungsseite –, scannt man sie, ohne sie zu zerschneiden, in mehreren Ablichtungen, vier oder sechs, die jeweils rund ein Viertel überlappen. Auf die Ränder braucht man dabei nicht zu achten. Diese Bilder werden der Panoramasoftware übergeben, die ja nicht nur einzelne Reihen von Fotos „zusammenkleben“ kann, sondern ganze Puzzles (Matrix-Anordnung). Man sagt ihr noch, dass es sich um abschnittsweise gescannte Bilder handelt, und schwupp ist die ganze Zeitungsseite aus vier oder sechs Scans zu einem Bild zusammengefügt.

Mit gängiger Bildbearbeitungssoftware können Korrekturen gemacht werden, Ränder beschnitten, Flecken getilgt, stö-

rende Anzeigen übermalt werden. Zum guten Ende wandle man noch in ein Schwarzweiß- oder Strichbild, falls man farbig fotografiert hat. Will man schließlich zum Speichern wirklich kleine, scharfe Dateien bekommen, so versuche man sich statt mit dem JPG-Format mit GIF, dem Graphics Interchange Format. JPG ist eher etwas für Buntes, GIF für Schwarzweiß. Nur das unkomprimierte Bitmap-Format, BMP, meide man ganz.

Und nach dem Scan, Tage, Jahre später? Wie findet man da den digitalisierten Mietvertrag oder das Foto von Vetter Franz wieder? Bilder sind inhaltlich noch nicht durchsuchbar. Desto wichtiger ist ein aussagekräftiger Name für die Datei, sei sie ein Scan oder ein Foto. Auch lassen sich beliebige Bemerkungen im Inneren von JPG-Bildern unterbringen („Metadaten“, „Exif-Daten“, „IPTC“, „XMP“). Etwas anderes ist ein Kommentar zur Datei, von Windows im Dateikopf abgelegt (Explorer, Eigenschaften, Dateinfo, Einfach). Solche, für alle Dateien mögliche Eintragungen werden im eigenen System vom Explorer und vor allem von Google-Desktop gefunden. Sie überleben aber – anders als wirklich im Bild eingebettete Daten – keinen Postversand.

Stehen Texte im Bild, etwa bei gescannten Zeitungsartikeln, wandelt man die JPG-Bilder mit Vorteil in PDF-Dateien um. Adobes Acrobat – der teure Writer, nicht der kostenlose Reader – hat eine Texterkennung, „Paper Capture“ genannt. Im Anschluss kann man in den PDF-Darstellungen nach Wörtern suchen. Praktisch beim Acrobat ist schließlich eine Optimierung, die die PDF-Dateigröße ohne merklichen Verlust an Qualität stark verringert. Immer wieder aber das Wichtigste: Suchprogramme wie Google-Desktop oder der gemeine Windows-Explorer (ab Windows XP SP2) durchsuchen PDF-Inhalte. Gibt man ein beliebiges Wort aus einem gescannten, in PDF gewandelten (und OCR-gelesenen) Artikel ein, so wird einem alsbald der Dateiname angezeigt. Die Erfindung der Schrift war eben immer noch die höchste Kulturleistung, jedenfalls für den PC. FRITZ JÖRN