

Bilder aus PDF-Dateien extrahieren

In Ausgabe 47 der ZPG-Mitteilungen wurde beschrieben, wie man in Word-Dateien eingebettete Bilder auf einfache Weise zur Weiterverwendung in ursprünglicher Qualität und Auflösung extrahieren kann. Da immer häufiger auch PDF-Dateien aus verschiedenen Quellen zur Verfügung stehen, werden hier einige Beispiele der Bildentnahme aus PDF-Dateien beschrieben.

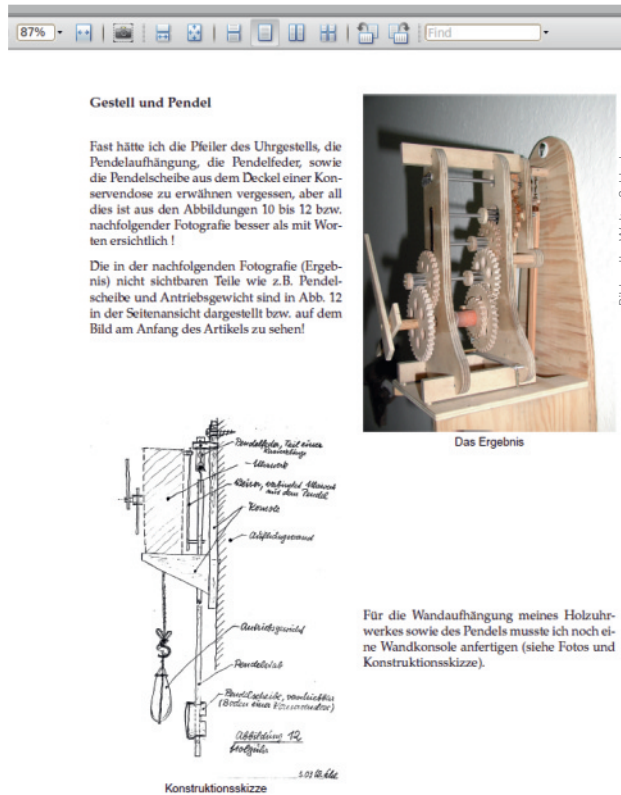
Grundsätzliches

Sowohl in Word- als auch in PDF-Dateien sind die kompletten Daten eingebetteter Bilder in verschlüsselter Form innerhalb des Textes enthalten. Ein separates Abspeichern der Bilder ist deshalb nicht ohne weiteres möglich. Selbstverständlich kann über das Anfertigen eines Screenshots prinzipiell jedes Bild „abfotografiert“ werden, allerdings nur in der momentanen Bildschirmauflösung, was nicht für jeden Zweck ein optimales Ergebnis liefert.

Ein Beispiel unter Windows

Mit dem Freeware-Programm *Some PDF Image Extract* können auf einfache Weise Bilder entnommen und im gewünschten Zielformat abgespeichert werden.

Die PDF-Datei mit eingefügten Bildern in verschiedenen Formaten



Bildquelle: Walter Schlenker

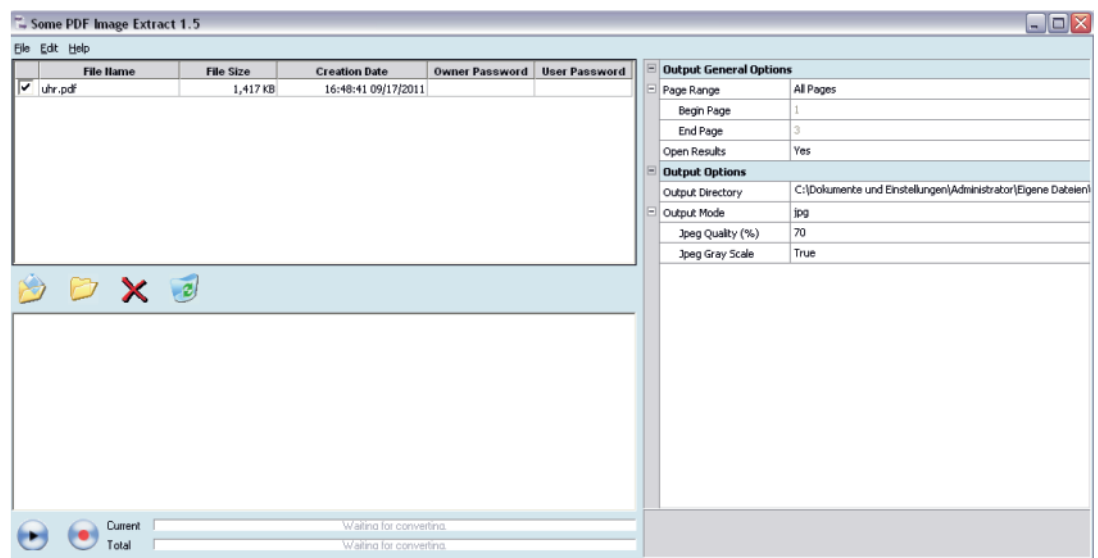
Aufgabe

Speichern Sie die Bilder einer gegebenen PDF-Datei in einem separaten Ordner

Vorgangsweise

- Laden Sie das Programm aus dem Internet herunter und installieren Sie es.
- Wählen Sie eine PDF-Datei aus mit FILE - OPEN > Sie wird oben links in der Tabelle angezeigt.
- Stellen Sie im rechten Fensterbereich nach einem Linksklick in ein weißes Feld die Ausgabeoptionen ein (z. B. ALL PAGES oder SELECTED PAGES, falls alle oder nur bestimmte Seiten ausgewählt werden sollen).
- Starten Sie die Konvertierung mit dem Startknopf links unten.
- Nach Eingabe eines (Teil-)Dateinamens (die laufenden Nummern werden vom Programm angehängt) startet die Bildentnahme. Der Fortschritt wird in den Feldern rechts neben dem roten „Panikknopf“ (links unten, zweiter Knopf) angezeigt.

Einstellung der Optionen für die Extrahierung



Screenshot: Some PDF (Freeware-Version)

Start auf Knopfdruck

Das Ergebnis

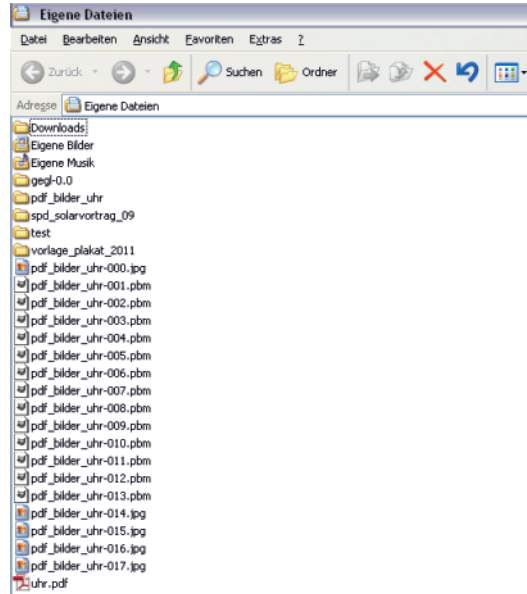
Die einzelnen Bilder befinden sich nun im ausgewählten Ordner.

In der Grundeinstellung von Windows werden allerdings die Dateieindungen wie z. B. *.jpg oder *.png usw. nicht angezeigt, so dass der Dateityp des Bildes nicht sofort festgestellt werden kann. Sofern noch nicht geschehen, sollte die Anzeige der Änderungen in den Einstellungen nachgeholt werden.

Dateieindungen anzeigen

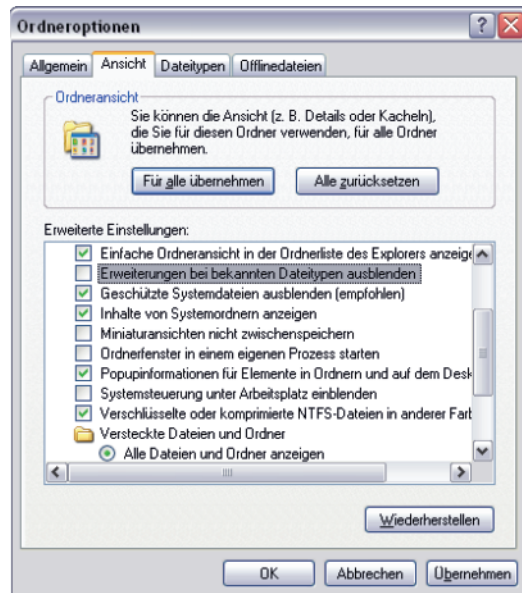
Dies geschieht in der Dateiübersicht mit EXTRAS - ORDNEROPTIONEN - ANSICHT durch das Entfernen des Häkchens vor ERWEITERUNGEN BEI BEKANNTEN DATEITYPEN AUSBLENDEN.

Tipp: Belassen Sie diese Einstellung, da sie auch für andere Dateien wie z. B. Texte, Kalkulationstabellen, PDF-Dokumente usw. die Übersicht verbessert!



Screenshot: Windows System

Die extrahierten Bilder mit Präfix und Nummerierung

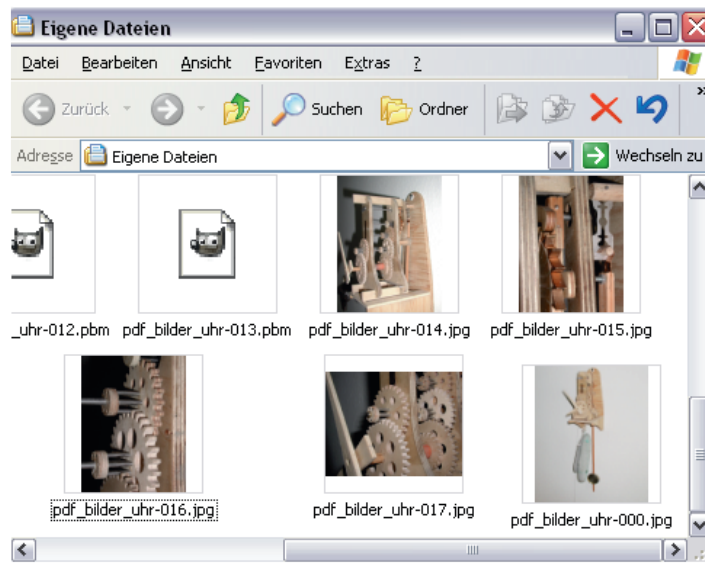


Screenshot: Windows System

Das Ausblenden der Dateieindungen sollte deaktiviert werden

Vorschau und Übersicht

Mit ANSICHT - MINIATURANSICHT kann eine kleine Vorschau der *.jpg-Bilder angezeigt werden. Die als *.pmb oder *.ppm extrahierten Grafiken (abhängig von der Bildquelle) werden allerdings nicht als Vorschau angezeigt. Sollten sich diese Bilder unter Windows nicht öffnen bzw. bearbeiten lassen, können Sie z. B. das Freeware-Programm GIMP installieren. Damit lassen sich die meisten Bildformate öffnen, bearbeiten und in allen gängigen Formaten wieder abspeichern.



Bildquelle: Walter Schlienger

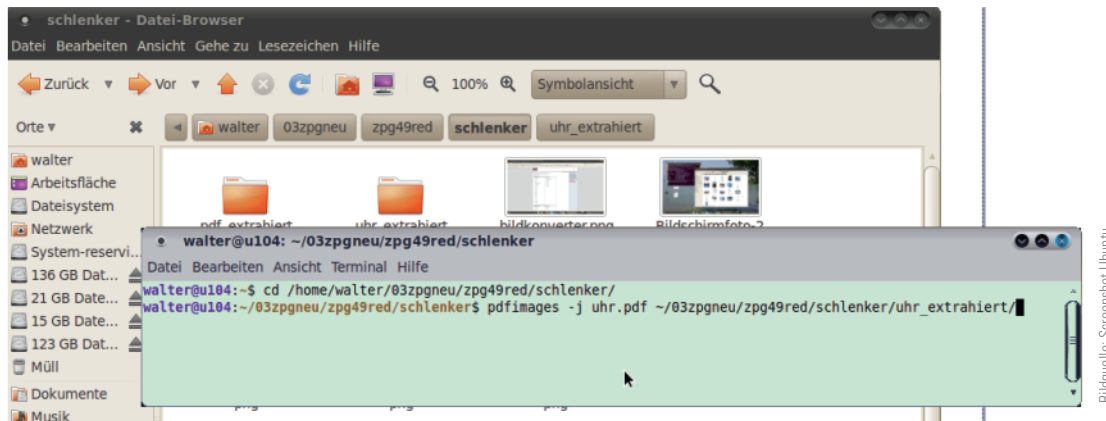
Die Bilder in der Vorschau

Bildentnahme unter Linux

Bei den meisten Linux-Distributionen sind die notwendigen Werkzeuge (*poppler-utils* und *imagemagick*) bereits vorinstalliert. Zur Sicherheit kann deren Vorhandensein nochmals über die Paketverwaltung überprüft werden, indem man nach den genannten Paketen suchen lässt.

Sind alle Voraussetzungen erfüllt, genügt ein kurzer Befehl in einem Terminal bzw. in der Kommandozeile (Aufruf durch Alt-F2), um das gewünschte Ergebnis zu erzielen:

Bildentnahme aus PDF-Dateien über die Befehlszeile unter Linux

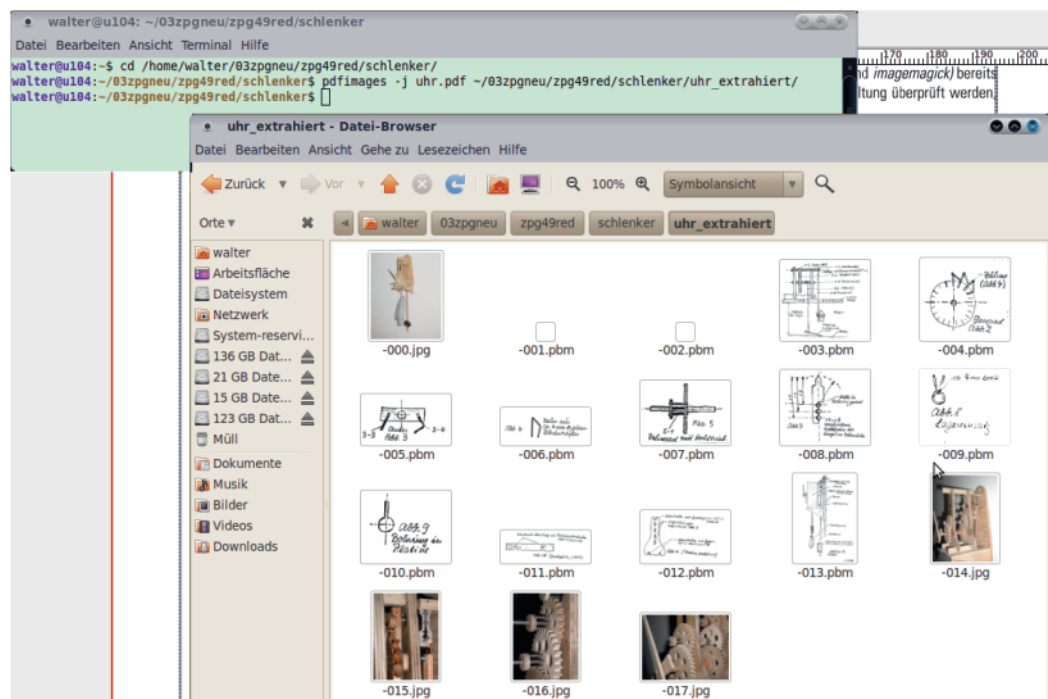


Bildquelle: Screenshot Ubuntu

- Rufen Sie ein Terminal (Befehlszeilenfenster) auf mit dem Menüpunkt ANWENDUNGEN - ZUBEHÖR - TERMINAL, falls noch kein Symbol dafür angelegt wurde.
- Wechseln Sie in das Verzeichnis mit der PDF-Datei (z. B. `datei.pdf`) mit `cd /home/ihr_name/verzeichnis_mit_der_pdf-datei`
- Geben Sie folgenden Befehl ein:
`pdfimages -j datei.pdf ~/verzeichnis_der_pdfbilder` [Enter] (wobei ~/ Ihr Home-Verzeichnis bedeutet und das Bildverzeichnis vorher angelegt werden sollte !)

Tipp: Verwenden Sie beim Eintippen die automatische Wortergänzung durch Drücken der Tabulatortaste. Manchmal genügt bereits ein Zeichen + [Tab], um den kompletten Verzeichnisnamen zu schreiben!

In der Symbolansicht werden unter Linux alle Bilder als Miniaturen angezeigt. Über die Lupensymbole können diese beliebig vergrößert (+) bzw. verkleinert werden (-).



Bildquelle: Screenshot Ubuntu

Konvertierung der Bildformate

Nicht alle Textverarbeitungs- oder Layoutprogramme können die erzeugten *.ppm- oder *.pbm-Dateiformate lesen und einbinden. Darüber hinaus sind z. B. *.ppm-Dateien sehr groß, da sie unkomprimierte Bilddaten enthalten (z. B. 2,5 MB statt 150 KB als *.png). Einzelne Bilder können z. B. mit GIMP geöffnet und anschließend im gewünschten Format gespeichert werden. Liegen dagegen mehrere Bilder im falschen Format vor, sollten diese zusammen in das gewünschte Format konvertiert werden. Auch hierzu gibt es zahlreiche Möglichkeiten unter den verschiedenen Betriebssystemen.

Unter *Windows* bietet sich z. B. IrfanView an, welches für Privatzwecke kostenlos erhältlich ist. Eine ausführliche Beschreibung des Programms erhalten Sie z. B. unter <http://www.filmscanner.info/IrfanView.html>.

Verwenden Sie *Linux* mit installiertem *Imagemagick*, genügt wieder ein kleiner Kommandozeilenbefehl, z. B.:

```
convert *.ppm bildname%d.jpg
```

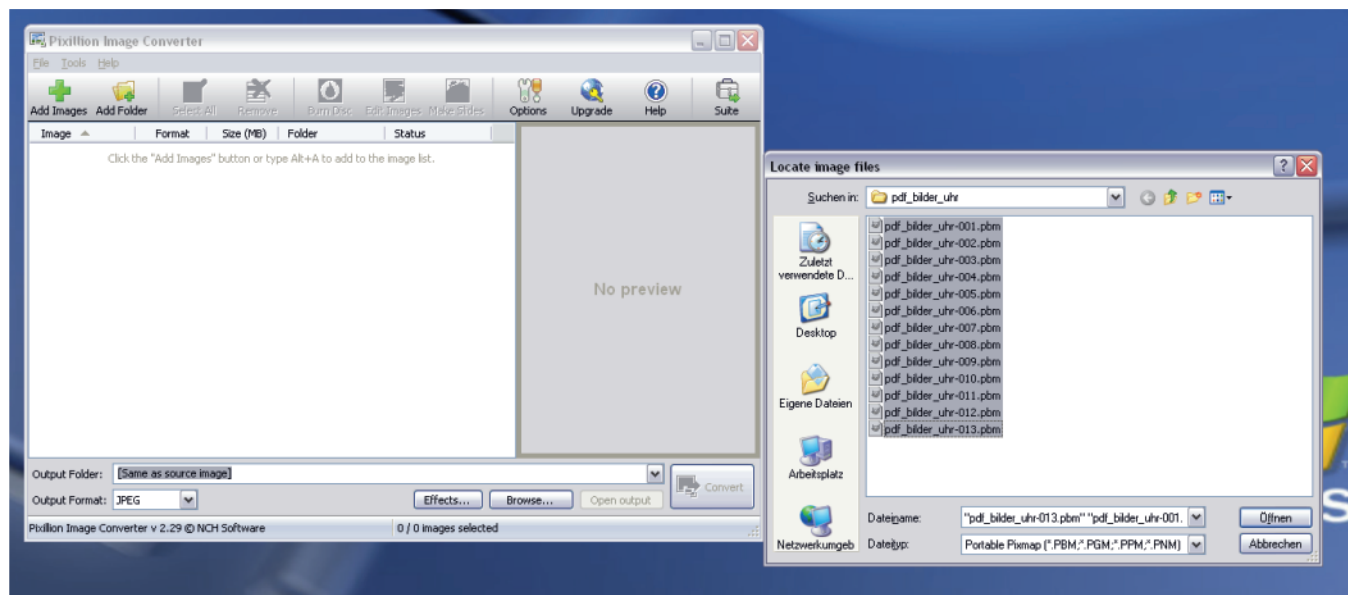
Dieser Befehl wandelt alle Bilder mit der Endung ppm in Bilder der Form BildnameBildnummer.jpg um. Bei 150 Bildern ergeben sich dadurch jpg-Bilder mit den Bezeichnungen Bildname000.jpg bis Bildname149.jpg. Die ursprünglichen Bilder bleiben erhalten.

Vor der Befehlseingabe müssen Sie im Terminal wieder in das Verzeichnis der zu konvertierenden Bilder wechseln.

Sowohl unter *Windows* als auch unter *Linux* (mit *Wine*) kann der *Pixillion Bilddatei-Konverter*, den es für den Privatgebrauch auch in einer Freeware-Version gibt, unter einer grafischen Benutzeroberfläche Bilder komfortabel konvertieren. (Download z. B. bei pixillion.softonic.de)

Pixillion

Mit Add Images werden die zu konvertierenden Bilder ausgewählt und anschließend mit dem Schalter Convert in das gewünschte Format umgewandelt.



Bildquelle: Screenshot Pixillion Freeware

Beachten Sie bitte, dass sämtliche hier erwähnten Programme weitaus mehr Funktionen als nur die Format-Konvertierung bieten. Diese zu beschreiben würde aber den Rahmen dieses Artikels bei weitem sprengen. Bitte informieren Sie sich ggf. auf den Homepages der Hersteller oder Anbieter.

Walter Schlenker