Gedanken zum papierlosen Büro

Kürzlich war in einer Computerzeitschrift ein Programm zum Thema „Papierloses Büro“ beigelegt. Da ich mich nicht gleich auf ein, wenn auch kostenloses Programm einlassen wollte, habe ich mir erst einmal Gedanken zu allen Notwendigkeiten gemacht um ggf. etwas simpleres selbst zu organisieren bzw. zu programmieren. Ein Hinweis auf einen, auch mobil nutzbaren, bestens geeigneten Scanner brachte mich zu weiteren Überlegungen.

Der Scanner ist der Canon P-150. Er scannt recht schnell bis zu 20 eingelegte Seiten und dies intelligent, d.h. er kann doppelseitig scannen und erkennt, ob eine „Rückseite“ vorhanden ist. Ebenso erkennt er die Formate Schwarz/Weiss oder Farbe. Das Speichern als PDF ist ebenfalls Voraussetzung.

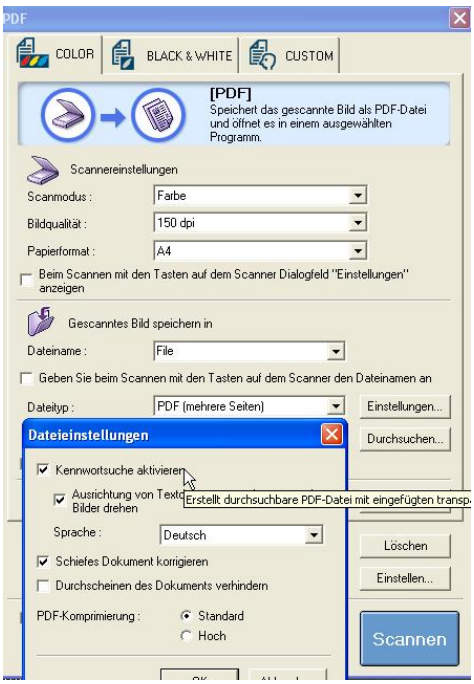
Im Vergleich zu früheren Zeiten lassen sich nun sehr wohl PDF- Formate inhaltlich durchsuchen und damit sind sie ideal zu Archivierungszwecken geeignet. Da es mittlerweile neben dem Urgestein Adobe noch viele weitere PDF- Ausgabeprogramme gibt, welche alle gewisse individuelle Eigenheiten besitzen, entstand die Notwendigkeit für eine Langzeitarchivierung, das PDF- Format zu Normen. Dies ist mit dem Format **PDF/A** erstmals geschehen und wird auch weiterhin gepflegt.

So weit die Theorie. Der P-150 hat auf der Rückseite einen Schiebeschalter mit dem man auf internen oder externen Modus umschalten kann. „Extern“ bedeutet, er verhält sich wie ein normaler Scanner der im Betriebssystem mit einem Treiber installiert sein muss. „Intern“ bedeutet, das im P-150 ein eingebauter „Speicherstick“ aktiviert wird, von welchem man ein Scanprogramm starten kann. Damit kann man den P-150 an einem beliebigen Windowsrechner betreiben. Der Vorteil des internen Betriebes hat jedoch eine Einschränkung - man kann beim Speichern als PDF- Datei im internen Modus keine Kennwortsuche durchführen (Volltextsuche erzeugen)! Da mit „Kennwort“ normalerweise Passwörter gemeint sind, ist diese Bezeichnung nicht glücklich gewählt. Besser ist „indizierter Text“. Dazu gehören Begriffe wie „Textsuche“ oder „Volltextrecherche“.

Dies funktioniert im „Externen Betrieb“. Auch mein stationärer „Canon 4400" beherrscht dies. Genutzt wird dazu ein, in der installierten Scannersoftware vorhandenes oder ein eigenständiges OCR-Programm.

Ein Scanner kann lediglich Bildpunkte (Pixel) erkennen und einlesen. Zur Texterkennung ist also eine Software erforderlich, welche eine Pixelgruppe einem Textzeichen zuordnen kann. Diese erkannten Textzeichen werden in einer PDF- Datei ggf. zusätzlich gespeichert und können somit auch bei einer „Textsuche“ gefunden werden.

Ansicht der „Textstellen“ einer PDF-Datei mit dem Editor „Notepad++“



Einstellungen des Canon 4400 für eine direkte Ausgabe als PDF -Datei mit aktivierter Kennwortsuche.

Hier kann sich das OCR nicht entscheiden ob „I“ oder „l“

Mein Canon 4400 nutzt das mitgelieferte OCR- Programm Omnipage SE4.0 zur Texterkennung. Da ich zudem nur mit einer „Genauigkeit“ von 150dpi gescannt habe ergeben sich einige Texterkennungsfehler.

Zwischenergebnisse:

1. Gescannte Belege und Dokumente belegen sehr viel Speicherplatz gegenüber einer direkten Erstellung aus einer Text- oder Tabellenbearbeitung heraus (s.u.).
2. Die Erkennung von Worten für die Volltextsuche in gescannten Dokumenten kann unzuverlässig sein!

PDF-Speicherbedarf Scannen vs. aus Dokument erstellen

Mit einer höheren dpi Zahl beim Scannen und einem aktuellen OCR- Programm sind allerdings gute Ergebnisse zu erzielen. Eine Garantie auf eine fehlerfreie Erkennung gibt es aber nicht.

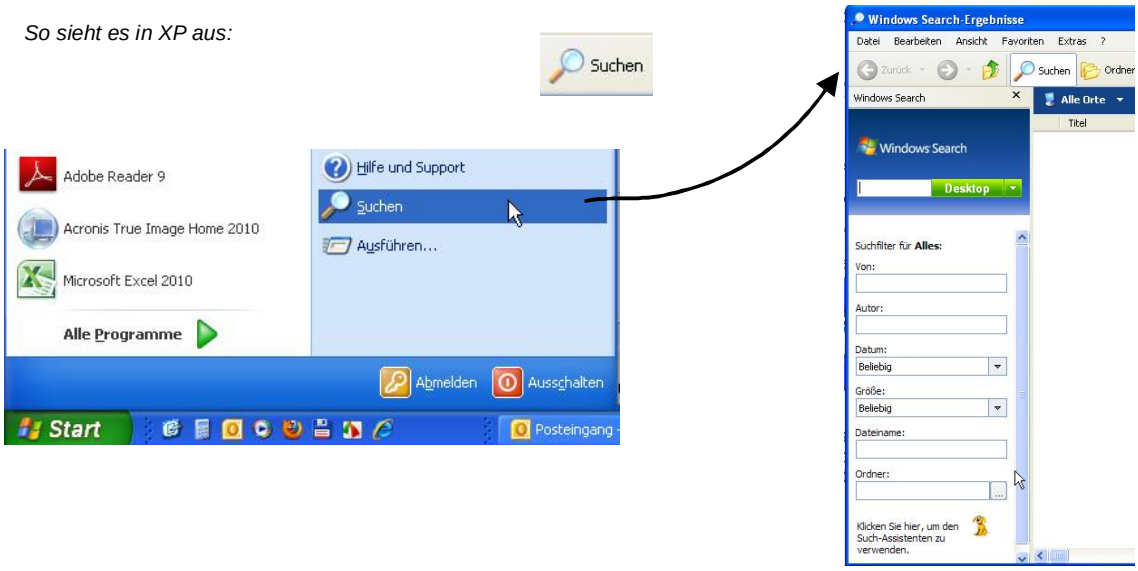
Mit dem Canon Scanner P-150 wird das Programm PaperPort in der Standardversion mit geliefert. Es hat vermutlich eine intergrierte OCR- Erkennung. Empfohlen wird jedoch ein höherwertiges Omnipage zu installieren (aktuell ist zur Zeit Version 18). Dieses wird dann für die Volltextsuche genutzt.

Es stellt sich nun die Frage ob ich nicht mit einfachen Mitteln meine Dokumente beherrschen kann?

Da mir jetzt klar war, dass PDF´s per Volltextsuche nach Begriffen durchsucht werden können fehlte nur noch ein Suchprogramm. Auch hier wurde mir erstmalig bewusst, dass schon lange entsprechendes Werkzeug auf meinem PC schlummerte. Das Programm „Windows Search“ (kurz WS) kann dies alles.

Seit Windows Vista, also auch in Windows 7 ist „Windows Search“ integrativer Bestandteil des Betriebssystems.

So sieht es in XP aus:



Windows Search (kurz WS) muss unter Windows XP optional geladen werden und steht auf der Mircrosoftseite als Download zur Verfügung. Siehe: Windows Search 4.0 für Windows XP (KB940157)

Das Betriebssystem Windows bietet als Grundfunktion für jede gespeicherte Datei vier Suchkennzeichen an. Diese Kennzeichen (auch Indexe) sind: Dateiname, Dateityp, Pfad und Änderungsdatum.

> Die begrenzten Möglichkeiten, diese Kennzeichen als Dokumentensuchfilter zu nutzen, ist ersichtlich. Mit „Windows Search“ kann man nun auch Dateiinhalte als Suchkriterium heranziehen.

Bevor man sich nun für ein „**D**okumenten**m**anagement- **S**ystem“ (DMS) entscheidet, kann man somit testen, ob taugliche Suchkriterien in den schon vorhandenen PDF´s vorhanden sind. Während man als Privatperson mit der Dokumentenverwaltung von „PaperPort“ in der Standardversion schon einiges bewerkstelligen kann, ist in betrieblicher Umgebung ein Profi- DMS unabdingbar.

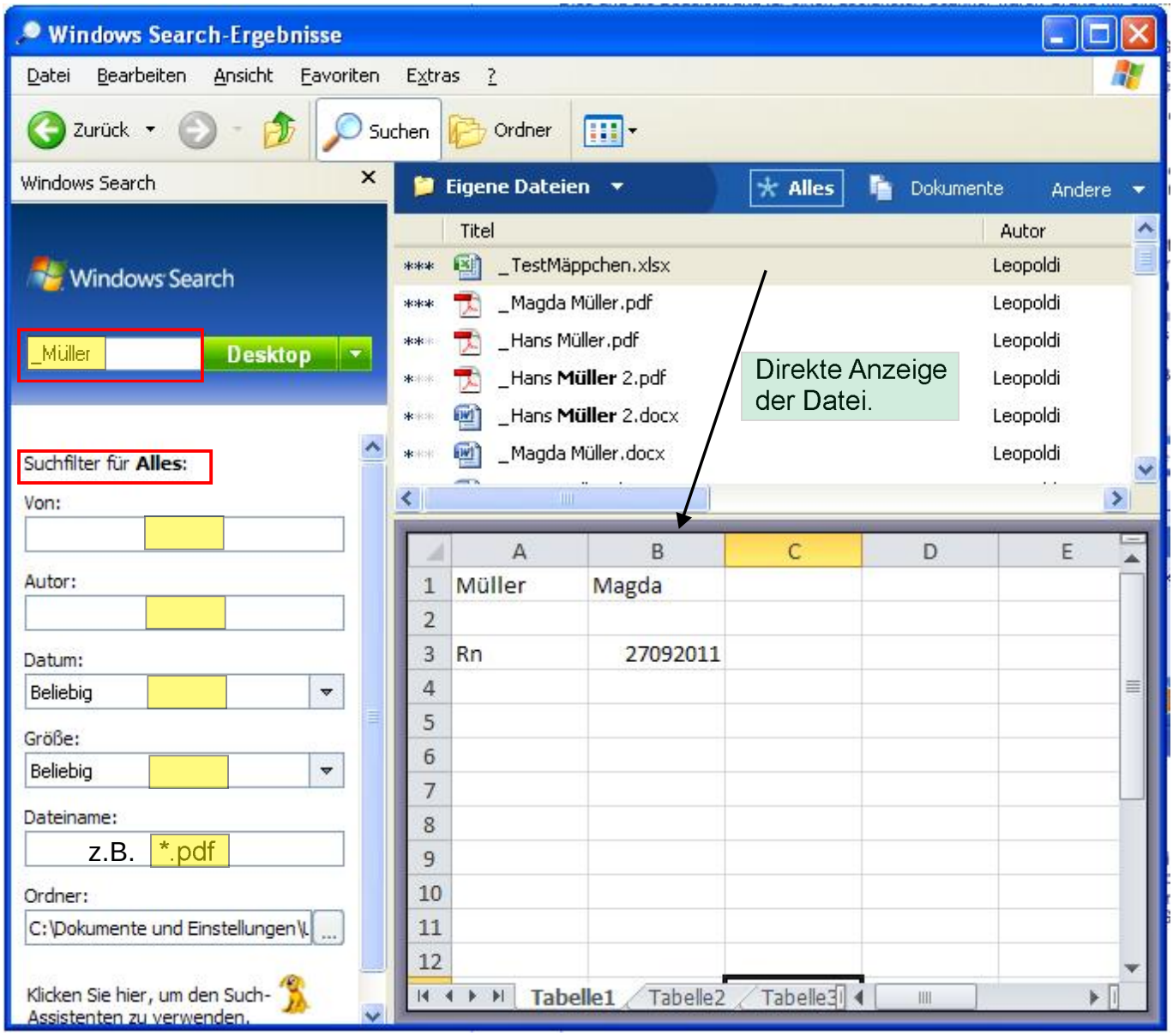
Aus „Die DMS-Fibel“ , Windream GmbH

Ein DMS erfüllt grundsätzlich mehrere Aufgaben: Die 7 V´s

- Verarbeiten, verwalten, vernichten, verwahren, verändern, verhindern und verfügen -

Beispielsuche mit „Windows Search“ in XP - Ausgeführt über START | Suchen

Zusätzlich zur Eingabe „_Müller“ im Suchfeld kann man über das Suchfilter sinnvolle Einschränkungen für die Suche vorgeben. Bei großen Datenmengen werden neben der Auswahlreduzierung zusätzlich Geschwindigkeitsvorteile erreicht.

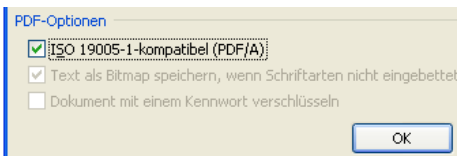


Mit Windows Search kann man also beliebige Textinhalte in Dateien suchen (Volltextsuche).

z.B.: Name, Anschrift, Rechnungs-, Kundennummer, Ort, Datum, Mahnung, Betreff, Zeichnungsnummer, Version u.v.m.

Auch technische Zeichnungen (CorelDraw, TurboCAD/ etc.) können, wenn sie als PDF mit Volltextsuche dokumentiert wurden, die oben genannten zusätzlichen Suchfunktionen erfüllen.

Alle zu archivierenden Dokumente sollte man im dokumentensicheren Format PDF/A ablegen. Dieses Format kann man z.B. im MS Office 2010 unter Dateityp „pdf“ | Optionen wählen.

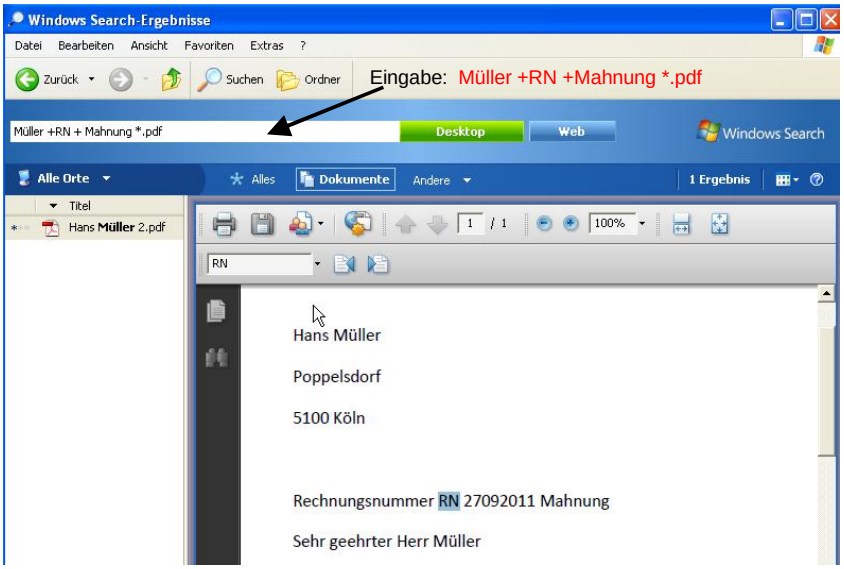


Generell kann man die Suche stark eingrenzen in dem man als Dateiname -Extension *.pdf eingibt.

Beispielsuche mit „Windows Search“ in XP beim Start aus der Taskleiste

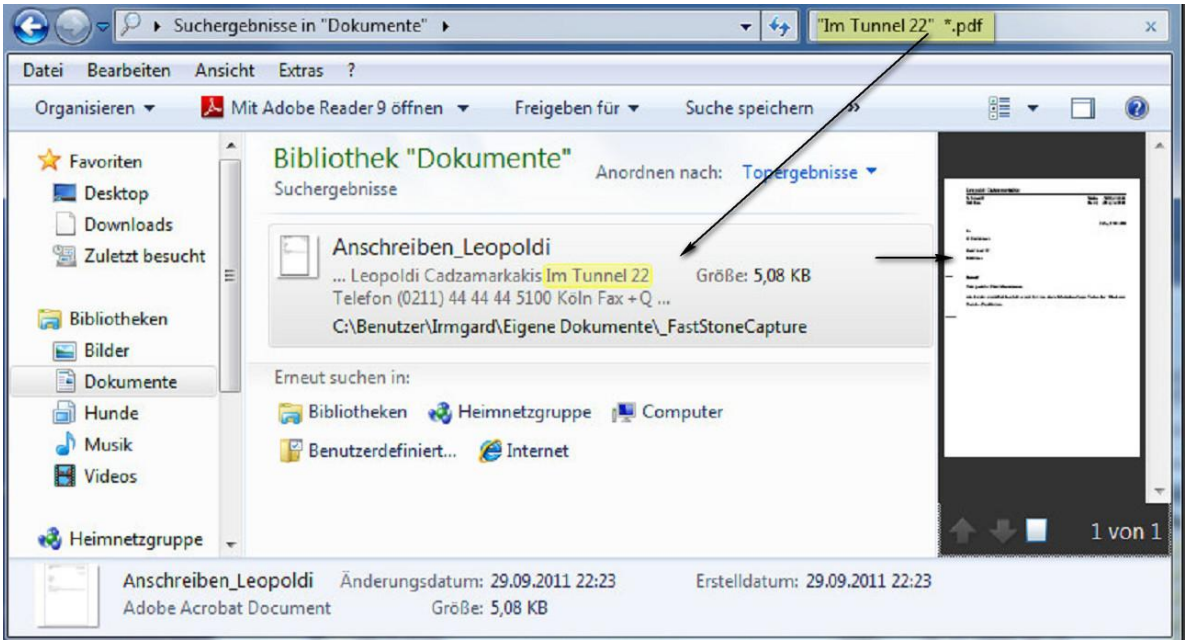
Mein „Windows Search“ in der Taskleiste erzeugt standardmäßig eine etwas andere Maske.

Hier sieht man wie man statt der Einstellungen im Suchfilter gleich bei der Suchtext-Eingabe selektieren kann.



In Windows 7 ist das integrierte „Windows Search“ informativer aber nicht mehr so übersichtlich.

Anm.: M.E. entwickelt sich das Betriebssystem Windows zu einer Oberfläche für Spielkinder und virtuelle Geister welche eine Spielkonsolenoberfläche benötigen. Wer denkt noch an den Büroalltag bei MS?



Auch PDF´s können kommentiert und Textstellen markiert werden. Ich nutze hierzu PDF-XChange View. PDFXCview gibt es neben der Pro-Version auch als eingeschränkte Freeware sowie als Portable-App für die mobile Nutzung auf einem USB- Stick..

Anm.: Mit PDFXCview kann man PDF-Formulare ausfüllen und speichern. Dies geht nicht mit dem Adobe Reader.

Meine Info´s habe ich aus folgenden Quellen, welche ich auch für weitere Informationen empfehlen möchte:

- Wikipedia; Suchbegriff „DMS“
- PaperPort von Nuance.de
- „Die DMS-Fibel“ , Windream GmbH

Alle meine Überlegungen basieren auf einer kurzen Recherche zu diesem Thema und sind nicht zwingend richtig oder gar vollständig!

Gruß Leopodi