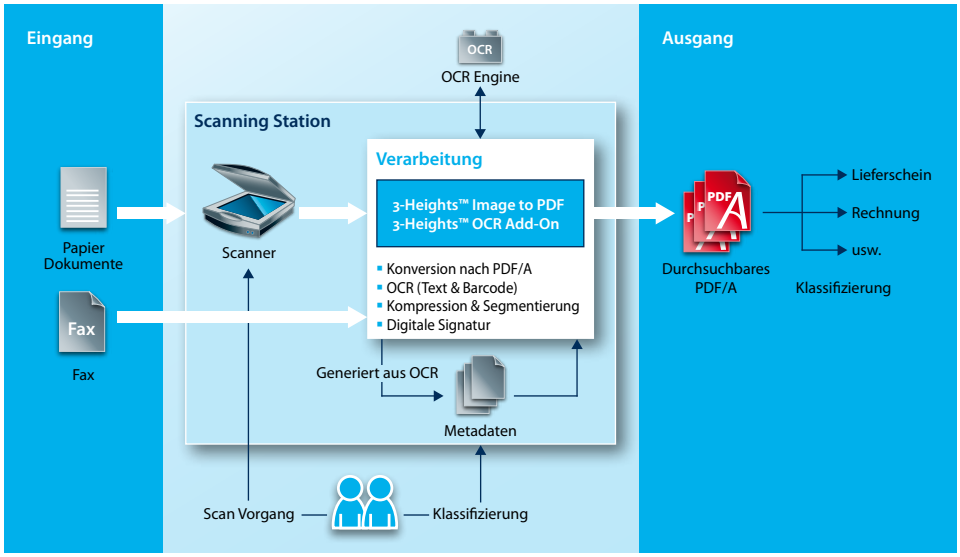
PDF-TOOLS.COM
Premium PDF Technology

Eine PDF/A-Lösung für gescannte Dokumente

Das Scannen von Papierdokumenten im Posteingangsbereich einer Unternehmung ist zum Alltag geworden. Oft wird diese Leistung von einem Scan-Dienstleister erbracht.
In den meisten Fällen werden die gescannten Bilder als TIFF-Dateien in Schwarz und Weiss erzeugt, so wie man dies von den FAX-Maschinen gewohnt ist. In speziellen Anwendungen wie Checks, Fotos für Ausweise usw. wird die Datei in Farbe erzeugt. Allerdings ist man damit sehr zurückhaltend, weil TIFF-Dateien in Farbe sehr gross werden können.
Der PDF/A-Standard hat sich heute in Posteingangs-Anwendungen, vor allem wenn es um das Scannen in Farbe geht, weitgehend durchgesetzt. Allerdings sind die einzelnen Bearbeitungsschritte wie Texterkennung, Kompression und Digitale Signatur in der Regel nicht optimal aufeinander abgestimmt und nicht in einer Lösung integriert. So gibt es beispielsweise Scanner, die bereits PDF/A-Dateien erzeugen und sie auch signieren können. Das nachträgliche Komprimieren bricht jedoch die Signatur und macht sie wertlos.

Der Lösungsansatz

Die PDF Tools AG bietet für das Erzeugen von PDF/A-Dateien aus gescannten und via FAX empfangenen Bildern eine Lösung an, welche die wichtigsten Anforderungen wie kleine Dateigrösse, Durchsuchbarkeit und eingebettete Metadaten erfüllt. Das folgende Bild zeigt das Prinzip.



Der Ablauf ist wie folgt:

1. **Bildakquisition:** Der Scan-Operator startet den Scanvorgang und erzeugt eine TIFF-Datei in Farbe. Der Scanner legt Dateien in der Regel in einem Dateiordner ab. Eine direkte Schnittstelle zur Scan-Software (via TWAIN) ist auch möglich. Faksimile-Dokumente werden von der FAX-Maschine empfangen und als

Vorteile der Lösung

- Skalierbarkeit mittels konfigurierbarer Optionen
- PDF/A-Dateien in Farbe mit kleinen Dateigrössen
- OCR-Erkennung möglich
- Einsatz Digitaler Signatur möglich
- Validierung der PDF/A Konformität

PDF Tools AG | Kasernenstrasse 1 | 8184 Bachenbülach | Switzerland
Tel.: +41 43 411 44 51 | Fax: +41 43 411 44 55 | pdfsales@pdf-tools.com | www.pdf-tools.com

TIFF-Dateien in Schwarz und Weiss in einen speziellen Ordner abgelegt.

2. **Manuelle Klassifikation:** Wahlweise und je nach Prozess führt der Scan-Operator eine Klassifikation durch, indem er den Scanner so steuert, dass die Bilder in verschiedenen Dateiordnern abgelegt werden z. B. für Rechnungen oder Lieferscheine.
3. **Segmentierung und Kompression:** Das Farbbild jeder Seite wird in seine Bestandteile wie Hintergrund, Text und Bilder zerlegt. Die einzelnen Teile werden durch spezifisch dafür entworfene Kompressionsverfahren in der Grösse reduziert. Dieses MRC-Verfahren (Mixed Raster Content) ermöglicht Farbdokumenten, konkurrenzfähige Dateigrössen zu erreichen.
4. **Texterkennung und Barcodes:** Die Bilder werden durch eine OCR-Maschine (Optical Character Recognition) weiterverarbeitet. Als Erstes wird das Bild entfleckt und gerade gerichtet, danach erfolgt die Erkennung des Textes und der Barcodes.
5. **Metadaten:** Informationen aus der manuellen Klassifizierung der erkannten Barcodes und weiteren Quellen werden zu standardisierten XMP-Metadaten zusammengefügt (Extensible Metadata Platform).
6. **PDF/A Erzeugung:** Die aufbereiteten Bilder jeder Seite, der erkannte Text und die Metadaten werden zusammen mit dem ICC-Farbprofil des Scanners zu einem PDF/A-Dokument zusammengefügt. Optional kann eine Index-Datei erzeugt werden, welche nur die Metadaten enthält.
7. **Digitale Signatur:** Wahlweise kann eine Digitale Signatur aufgebracht werden, damit die Nachvollziehbarkeit und Revisionsfestigkeit des Dokuments sichergestellt werden kann.
8. **Validierung:** Wahlweise können die PDF/A-Konformität des erstellten Dokumentes und die Gültigkeit der Digitalen Signatur überprüft werden.

Vorteile der Lösung

- Die Lösung kann von der einfachsten bis zur voll ausgebauten Form skaliert werden. Dafür gibt es die konfigurierbaren Optionen MRC, OCR und Digitale Signatur.
- Das MRC-Verfahren ermöglicht das Erzeugen von PDF/A-Dateien in Farben, die vergleichbare Grössen zu TIFF-Dateien in Schwarz und Weiss erreichen (in der Grössenordnung 40 kByte).
- Als OCR-Maschine kann ABBYY FineReader eingesetzt werden. Für einfachere Anwendungen kann alternativ die kostenlose OCR-Erkennung von Tesseract verwendet werden.
- Für das Aufbringen einer grossen Anzahl Digitaler Signaturen kann auch eine HSM (Hardware Security Module) von SafeNet eingesetzt werden.
- Die PDF/A-Konformität der erzeugten Datei kann mit einem Validierungsprotokoll nachgewiesen werden.

Installation und Betrieb

Der Scan-Server von PDF Tools AG kann als Dienst auf einem Rechner mit Windows-Betriebssystem verwendet werden.
Bestimmte Prozessschritte wie die OCR-Erkennung und das Aufbringen einer Digitalen Signatur können getrennt und auf verschiedenen Rechnern durchgeführt werden.
Komplexere Architekturen, wie die Texterkennung und das Einfügen von Metadaten nach dem Signieren, mehrfache Digitale Signaturen usw. sind im Rahmen eines konkreten Projektes möglich.

PDF-TOOLS.COM
Premium PDF Technology

Über PDF Tools AG
PDF Tools AG ist ein weltweit führender Hersteller von Softwarelösungen und Programmierkomponenten für die PDF und PDF/A Erzeugung, Bearbeitung, Wiedergabe und Archivierung. Die Architektur der Software wurde speziell auf die leistungsfähige Bearbeitung grosser Dokumentenvolumen ausgelegt. Die vielseitigen Produkte sind besonders für die Unterstützung von Arbeitsprozessen in Unternehmungen sowie als Komponenten für Integratoren und OEM-Kunden geeignet. Weitere Informationen finden Sie unter www.pdf-tools.com.

PDF Tools AG | Kasernenstrasse 1 | 8184 Bachenbülach | Switzerland
Tel.: +41 43 411 44 51 | Fax: +41 43 411 44 55 | pdfsales@pdf-tools.com | www.pdf-tools.com