

Akten(-Ordner) verpixelt euch!

Smart Capture, Dokumentenprozesse, Dokumentenformate, PDF/A-Standard, Scannen, OCR (Optical Character Recognition), Archivierung



Nadine Schuppisser ist Leiterin Marketing und Kommunikation der **PDF Tools AG**, einer international erfolgreichen Softwareentwicklungs- und Vertriebsgesellschaft. Das Unternehmen bietet PDF & PDF/A Komponenten und Lösungen rund um die Digitalisierung, den Dokumentenprozess und die gesetzlich konforme Langzeitarchivierung an. PDF Tools AG ist die Schweizer Vertretung im ISO Komitee für PDF/A und PDF und Gründungsmitglied der PDF Association.

www.pdf-tools.com

Digitalisierung scheint aktuell eines der meist geläufigen Wörter zu sein. Kaum ist ein Pixel involviert rückt das allseits beliebte Buzzword hinterher. Knapp dahinter purzeln viele Namen von Firmen, Digital-Philosophen und Datengurus. Doch was genau haben Firmen im Fall von Capturing davon?

Vergangenheit und (noch) Gegenwart

Weit vor der Digitalisierung wurde die Brief-Post geöffnet, sortiert, Eingangsstempel drauf und ab ins interne Postfach der zuständigen Person. Diese hat das Dokument gesichtet, zum Beispiel eine Rechnung kontiert und ins Buchhaltungssystem eingegeben. Danach wurde das Dokument zur Signatur an den Kostenstellen-Inhaber weitergereicht.

Nach der Freigabe kam die Rechnung wieder zurück, die Zahlung wurde ausgeführt und in einem dicken Aktenordner abgelegt, der später in einem Archiv im Keller landete. Archivschränke und Aktenordner in Kellerräumen wurden bis unter die Decke gefüllt, und wenn ein Beleg gefragt war, begann die Suche durch die Regale, Ordner und Register. Papier ist zeit- und kostenintensiver. Und wehe es hat jemand eine Akte oder einen Beleg falsch abgelegt! Noch stehen viele Firmen mit einem Fuss in der Vergangenheit – befassen sich aber aktuell intensiv mit der Digitalisierung.



Digitalisierung beim Capturing?

Bei Unternehmen können viele Prozesse durch die Digitalisierung einen Mehrwert bringen. Picken wir ein Buzzword raus: Capturing. Tagtäglich fluten uns Dokumente, Daten, E-Mails und weitere Informationen. Wir erhalten geschäftsrelevante und weniger relevante Informationen über diverse Kanäle in den verschiedensten Formaten. Digitale Prozesse können den Eingang dieser Daten überwachen, scannen, sortieren, kategorisieren und in die entsprechenden Arbeitsschritte fließen lassen. Die Digitalisierung soll als Unterstützung dienen, anstehende Prozessschritte optimal vorzubereiten. Dennoch sind Maschinen und Algorithmen nicht immer in der Lage, mit allen Fällen umzugehen.

Um überhaupt in Fahrt zu kommen müssen Eingangsdokumente gescannt werden, mit einer OCR-Engine die nötigen Informationen erkannt, rausgelesen und mit dem Dokument verknüpft werden. Nur so können zum Beispiel Rechnungen von Lieferanten in unterschiedlichsten Formaten erkannt und gelesen werden, damit die Unterstützung der elektronischen Zahlung auch funktioniert.

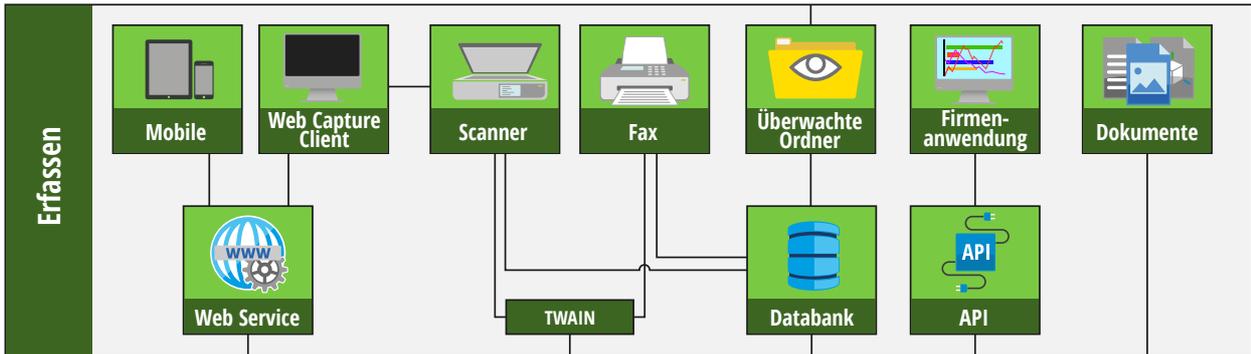
Durch die Texterkennung (OCR) erlangt man nicht nur einen schnelleren Durchlauf vom Eingang über die Kontierung bis hin zur Ausführung der Zahlung – durch die hinterlegten Informationen ist es auch jederzeit möglich, mit Suchbegriffen die

entsprechende Rechnung wiederzufinden. Die Papierform der Rechnung ist vom Tisch sobald diese gescannt, alle benötigten Daten hinterlegt, die Freigabe und Zahlung erfolgt ist und diese im elektronischen Archiv gesichert ist. Damit ist nicht nur das Abspeichern gemeint, sondern auch der Schutz des Dokumentes als solches bezüglich Zugriff, Zeitstempeln und elektronischen Signaturen.

Selbst der Arbeitsplatz eines jeden ist im Wandel. Durch diese Veränderung ist Flexibilität gefragt. Es gibt immer mehr Mitarbeiter, welche von zuhause aus arbeiten, aber die Dokumenteneinsicht benötigen. Durch das Digitalisieren kann der Zugang an die benötigten Informationen sichergestellt werden. Übrigens auch ein Traum für jeden „Clean-Desk“-Prüfer, welcher je länger je weniger auf den Tischen finden sollte. Daher gibt es auch Dokumenten-Prozesse, die wieder aus dem Unternehmen führen. Auch da ist die elektronische Form in den meisten Fällen deutlich schneller, sicherer und effizienter.

Es kommt noch besser! Bei Revisionen muss man nicht mehr kistenweise Aktenordner aus dem Keller in einen Besprechungsraum für den Revisor befördern – man gibt diesem einfach entsprechende Rechte für die Akteneinsicht, um im digitalen Archiv die Dokumente remote aufzurufen und zu prüfen. Das ist für beide Seiten zeit- und nervensparend. Wer das nicht möchte, kann die Dokumente auf einen gesicherten USB-Stick speichern und dem Revisor übergeben.





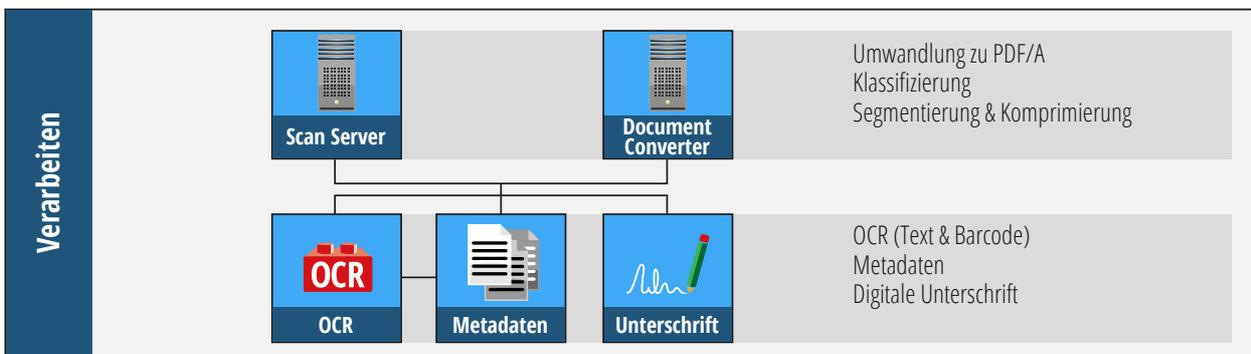
Verpixeleien – der digitale Kescher

Eingangskanäle für Dokumente sind zahlreich wie auch das Spektrum an Formaten und Qualitäten. Es wird „gecaptured“ wo es nur geht in der analogen wie auch in der digitalen Welt. Selbst das Papier hat im Zeitalter von E-Rechnung, Online-Schalter und E-Commerce keineswegs ausgedient: Dokumente wie Rechnungen, Steuerformulare, Service-Berichte und Verträge werden nach wie vor oft auf Papier ausgefertigt, per Post verschickt und auf dem Postweg entgegengenommen.

Um die Papierdokumente in elektronische Form für den Dokumentenprozess zu bringen, werden diese gescannt. Dabei ist gute Bildqualität wichtige Voraussetzung für eine gute Texterkennungsrage. Für eine hohe Kompressionsrate bei gleichzeitig guter Bildqualität benötigt man Rechenleistung, die in

den dezentralen Multifunktionsgeräten selbst oft nicht zur Verfügung steht. Traditionell erzeugt ein Scanner für jede Seite ein TIFF- oder JPEG-Bild. Einige von ihnen können direkt PDF-Dateien erzeugen. Und neuere Geräte produzieren Dateien, die dem PDF/A-Standard entsprechen. Die Qualität der produzierten Dateien ist jedoch je nach Anbieter, respektive Konvertierungssoftware sehr unterschiedlich.

Um für dieses Qualitätsproblem Abhilfe zu schaffen, empfiehlt sich ein zentraler Scan-Server. Der Dienst übernimmt alle Aufgaben, welche von den dezentralen Scan-Stationen an ihn delegiert werden können. Er nimmt die eingescannten Bilddateien entgegen, analysiert die Dokumente und erzeugt ein PDF/A-Dokument mit allen Text- und Bildinformationen in jeweils passender Kompression. Optional markiert er das Dokument mit einem Zeitstempel oder einer digitalen Signatur. Die erfassten Informationen stehen damit sowohl für



menschliche Leser als auch zur automatisierten Weiterverarbeitung mit IT-Anwendungen in standardisierter, qualitativ hochwertiger Form zur Verfügung.

Gegenüber steht die digitale Fraktion, deren Hersteller von elektronischen Formaten oft Dokumente in unzureichender PDF-Qualität produzieren, was zu unerwarteten Problemen und Kosten in der Dokumentenverarbeitung führt. Bei der Konvertierung in das PDF-Format geht es nicht nur darum, ein Bild in einen PDF-Umschlag einzubetten. Das PDF-Dokument kann Text- und Barcode-Erkennung, Einbettung von Metadaten und digitale Signaturen beinhalten.

Bedingt durch den rasanten technologischen Fortschritt, veralten heute eingesetzte Systeme rasch und werden eines Tages ersetzt. Die Inhalte der archivierten Dokumente hingegen bleiben relevant und müssen deswegen unverändert in die neuen Systeme migriert werden können. Wegbereiter für eine verlustfreie Migration ist ein stabiles, den Lebenszyklus der Systeme überdauerndes Dokumentenformat.

Vom Dot zum Bit

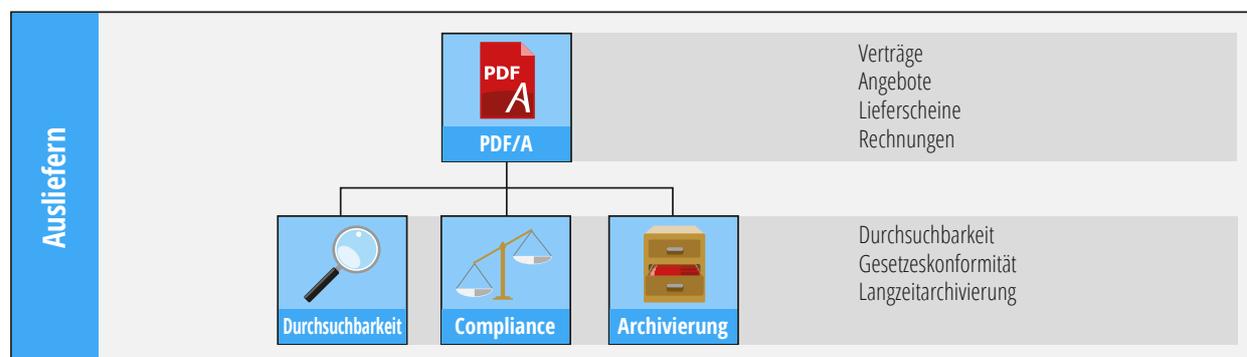
Erweitert man die Scanfunktion mit Texterkennung (OCR) erhält man die Option, sämtliche Dokumente mit Zusatzinformationen anzureichern, welche sich auf den gesamten Dokumentenprozess positiv auswirken. Vom gescannten Papier, zum elektronischen Dokument bis hin zur E-Mail mit Anhängen können die Dokumente aufgrund von Textmerkmalen, Elementen

und Dokumentaufbau Fachabteilungen oder Prozessen zugeordnet werden.

Dabei wird erkannt, ob es sich bei den Dokumenten beispielsweise um normale Korrespondenz, Bestellungen, Lieferscheine oder Rechnungen handelt und entsprechend dem Verteilerschlüssel weitergereicht. Dabei werden auch nachgelagerte Bearbeitungsprozesse und Entscheidungszyklen maßgeblich unterstützt.

Eine gute Texterkennung mit nachgelagerter Verarbeitung unterstützt somit nicht nur das Capturing als solches. Auch die Klassifizierung, Dokumentenverteilung inklusive deren Weiterverarbeitung sowie die bessere Überprüf- und Auffindbarkeit der Dokumente für Compliance-Zwecke können so effizient gestaltet werden. OCR ermöglicht nach der Informationsanreicherung das Suchen nach Stichworten, Reporting-Funktionen, das Zusammenstellen von themenspezifischen Dossiers und damit generell eine schnelle Reaktionszeit und Verarbeitung.

Auch wenn der Austausch von elektronischen Dokumenten in Geschäftsprozessen mittlerweile zur Selbstverständlichkeit geworden ist – wird der Blick auf die Dokumentenqualität oft vernachlässigt. In beiden Fällen ist ein Qualitätskontrollsystem für eingehende Dokumente unerlässlich geworden. Bevor man jedoch ein – nennen wir es „Quality Gate“ – einrichtet, muss man sich im Klaren sein, was Qualität bedeutet. Ein Großteil der Unternehmen richtet den Dokumentenprozess darauf aus, mit einem Hauptformat zu arbeiten. Dies erleichtert den gesamten Ablauf und vermindert den Format-Dschungel in den Kanälen und lässt sich somit deutlich effizienter kontrollieren. ▶



Nehmen wir PDF als Standardformat, so würden sich die zwei Arten von Qualitäten ergeben. Die inhärente Qualität und die dedizierte Qualität. Die inhärente Qualität prüft die Konformität mit der Dateiformatspezifikation (ISO 32000) – denn nicht überall wo PDF draufsteht ist auch PDF drin. Zumindest nicht so, wie es Unternehmen für die Weiterverarbeitung bis hin zum Archiv wünschen.

Die dedizierte Qualität konzentriert sich auf die Anwendungsfälle wie zum Beispiel Scannen, Dokumentenaustausch, Drucken und Archivieren. Beispielsweise ob Schriftarten und Farben optimiert sind und die Dateien bei Bedarf dem PDF/A-Standard entsprechen, wenn es um die Ablage im digitalen Langzeitarchiv geht. Ein „Quality-Gate“ sollte sich nach dem Capturing um die Validierung, Reparatur, Optimierung und die Digitale Signatur und somit den Schutz der Dokumente kümmern.

Wichtig ist, dass ein Dokumentenprozess – egal wie einfach oder umfangreich dieser sein mag – vor der Inbetriebnahme gut geplant und geprüft wurde. Sei es bezüglich des Umgangs mit externen und internen Dokumenten wie auch deren Prüfung, Anreicherung, Verteilung und Ablage. Dies verbessert die Reaktionszeit, schont Ressourcen auf allen Ebenen und ist deutlich transparenter für die Nachvollziehbarkeit und Compliance.

Fazit

Ein zentraler Scan-Server mit OCR (Optical Character Recognition) und ein nachgelagertes „Quality Gate“ sind Schlüssel zu einem sauberen digitalen Dokumentenprozess in Bezug auf Capturing. Die Texterkennung ermöglicht die gewonnenen

Informationen, die Wege eines Dokumentes zu kanalisieren, Dokumentenprozesse effizient zu gestalten und eine gute Durchsuchbarkeit. Das „Quality Gate“ sorgt für durchgehende Qualität und die Sicherheit der Dokumente. Durch die Automatisierung lassen sich Fehlerquellen minimieren, die Dokumentenqualität anheben und somit auch Zeit und Kosten sparen – und dies vom Eingang bis hin zur Ablage im Langzeitarchiv.

Neben den Bits und Bytes sollte der Faktor Mensch ebenfalls nicht vergessen werden. Sei es beispielsweise bezüglich der Benutzerfreundlichkeit oder dem Datenschutz. Man sollte beachten, dass eine technisch brillante Software selbst weder Akzeptanz bei Anwendern noch Transparenz hinsichtlich gesetzlicher Vorschriften und Richtlinien garantiert. Voraussetzung dafür sind ein gutes Konzept und Verständnis. Die Nachvollziehbarkeit ist nicht nur bei den digitalisierten Dossiers im Archiv relevant, sondern auch in den Köpfen der Menschen.

Capturing ist keine Magie, sondern schlichte Erfassung, Prüfung und Anreicherung von Informationen. Aber mit Capturing kann man viel erreichen – also lasst die Aktenordner sich verpixeln und sagt hallo zur Digitalisierung. ■

