

KGL Kopier Gesellschaft mbH
Floßplatz 24, 04107 Leipzig



postStream
DOKUMENTEN
MANAGEMENT

postStream

White paper

Januar 2016

Kategorie: Tipps und Tricks

Schreiben für Maschinen

Zusammenfassung

Moderne DMS und ECM Software hat den Büroalltag grundlegend verändert. Bürokratische Prozesse sind soweit automatisiert, dass sie nicht länger die Manpower eines Unternehmens binden und von den Mitarbeitern im Bestfall als Hintergrundprozesse wahrgenommen werden. Mit der Verschiebung der Aufmerksamkeit hat sich auch die Rolle von Schriftstücken geändert. Ursprünglich für menschliche Augen geschrieben, sind es heute Maschinen denen die Erstverarbeitung obliegt. Seien es Scanner, die papiergebundene Dokumente digitalisieren oder Programme für die optische Zeichenerkennung, welche digitale Dateien weiterverarbeiten. Bei der Erstellung eigener Unterlagen werden diese heute so wichtigen Komponenten der Dokumentenverarbeitung nach wie vor häufig ignoriert. An dieser Stelle begangene Fehler sind nicht nur imageschädigend, sie sind auch leicht vermeidbar. Dieses White Paper enthält eine Zusammenstellung regelmäßig begangener Fehler, die allesamt aus dem praktischen Alltag der Bestandskunden von postStream entnommen sind. Bei der Erstellung der Hinweise greifen wir auf einen Fundus von 25 Jahren Erfahrung zurück.

Inhaltsverzeichnis

Schreiben für Maschinen	3
Die neue Rolle von Papier	3
Der automatische Vorlagenwechsler	5
Digitale Texterkennung – OCR.....	10
Header	12
Fließtext	14

Schreiben für Maschinen

Leitfaden

Immer mehr Unternehmen verwenden heute Software für Ihre Dokumentenverarbeitung und könnten theoretisch komplett auf Papier verzichten. Aktuelle Studien zufolge nimmt der Papierverbrauch in Deutschland seit 2008 entgegen den Erwartungen wieder zu. Parallel zu dieser Entwicklung wächst aber auch der Anspruch an die Erstellung Ihrer Schreiben.

Die neue Rolle von Papier

Auch wenn heute sehr viele Prozesse wesentlich effizienter, komplett digital ablaufen könnten, sind die Wenigsten dazu bereit, auf händisches Material zu verzichten. Die rasante technische Entwicklung hat vielmehr dazu geführt, dass dem Papier innerhalb der Kommunikation zwischen Unternehmen eine erweiterte Rolle zukommt. Längst nicht mehr das sicherste Medium, was Informationserfassung und Speicherung betrifft, ist es zu einem Garant von Sicherheit geworden. Es erfüllt zunehmend die symbolische Funktion eines Händedrucks. Es setzt sich also an die Stelle jener gestischen Handlung, die beim Erstkontakt seinem Gegenüber zu verstehen gibt, dass man keine Waffen bei sich trägt. Art und Stärke des Händedrucks enthalten darüber hinaus wertvolle Informationen, welche vergleichbar auch beim Papier gefunden werden. Glattes Hochglanzpapier transportiert einen anderen Eindruck als Leporellopapier. Ein guter, kräftiger Händedruck würde etwa bei 100g/m² Papiergewicht liegen und so weiter. Das Layout wiederum spiegelt den Augenkontakt während des Händedrucks wieder. Es gibt wohl kein Unternehmen, welches seine ausgehenden Schreiben nicht mit einer Maske versieht, sprich individualisiert. Ein gutes Layout gleicht einer angenehmen Erscheinung, man schaut gerne zurück und zu Vertrauen gesellt sich Sympathie. Diese neue Bedeutung von Papier wurde unlängst erkannt und im Internet findet man zahlreiche Hilfen, wie man seine Schreiben, auf diese Aspekte bezogen, aufpöppelt. Doch vergessen diese Hilfen leider alle eine Kleinigkeit, die sich unter

Umständen fatal auswirken kann. Menschliche Augen und Hände kommen erst an zweiter Stelle mit den meisten Schreiben in Kontakt.

Den Erstkontakt haben Maschinen. Die automatischen Vorlagenwechsler¹ sind die neuen Hände und die optische Zeichenerkennung² die neuen Augen. Bevor also das Papier seine Wirkung beim Menschen entfalten kann, gilt es diese Hürde zu meistern. Leider fallen hier viele Unternehmen nach wie vor negativ auf. Wenn wir Nutzer von postStream als Referenz nehmen, so kann man davon ausgehen, dass ca. 30% aller Schreiben³ nicht auf digitale Erkennung hin optimiert sind. Solche Schreiben verzögern nicht nur Arbeitsprozesse und erzeugen Frust bei Mitarbeitern. Sie werfen ganz allgemein ein negatives Licht auf das Image des Unternehmens. Dabei bedarf es keines großen Aufwandes hier Abhilfe zu schaffen.

¹ In der Folge ADF genannt (deutsch AVW).

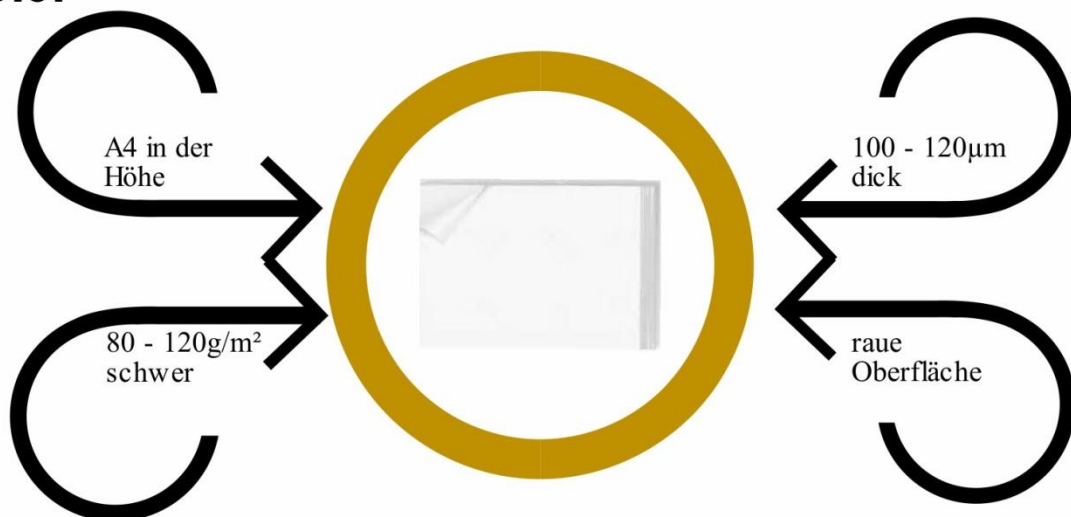
² In der Folge OCR genannt.

³ Nach Auswertung von 1000 papiergebundenen Schriftstücken lag der genaue Wert bei 31,75%. Es wurden sowohl Fehler im Zusammenhang mit den automatischen Vorlagenwechslern berücksichtigt, als auch solche die im Zusammenhang mit einer OCR entstanden. Je nach Schwere des Fehlers, die ein Schriftstück verursacht hat, wurde diesem ein Wert von 0; 0,5 oder 1,0 zugeordnet. Ein Ergebnis von 50% enthält also auch folgende mögliche Interpretation: „Alle Schriftstücke enthalten einen leichten Fehler.“. Der Großteil unserer Kunden verwendet Scanner von RICOH.

Der automatische Vorlagenwechsler

Unternehmen, die Software für die Dokumentenverwaltung nutzen, scannen in der Regel die gesamte eingehende Post. Dabei spielt die Stapelverarbeitung die wichtigste Rolle. Je nach Preisklasse können ADFs ein Papiervolumen von bis zu 80 Blatt pro Minute bewältigen. Hält man die Spezifikationen der Hersteller ein, ist die Papierstaugefahr minimal. Doch Spezifikationen mögen bei der Kaufentscheidung noch eine Rolle gespielt haben, im täglichen Gebrauch aber wird diese kaum ein Mitarbeiter kennen. Es schadet daher nicht, die eigenen Schreiben von Anfang an so anzufertigen, dass diese nach Möglichkeit die Durchschnittskriterien der handelsüblichen ADFs erfüllen. Die hier vorgestellten Hinweise zielen darüber hinaus auf die optimale Arbeitsleistung von ADF-Modulen, ein innerhalb der Spezifikationen eng begrenzter Bereich.

Papier



Papiergewicht:

Das Papiergewicht ist die Standardangabe für die Qualifikation von Papier allgemein und wird in g/m² angegeben.

Grundsätzlich gilt:

„Je schwerer das Papier, desto höher die Qualität, desto

bedeutender die Information, die dieses Enthält“. Gute ADFs geben in ihren Spezifikationen eine Mindestbetriebsreichweite von 60g/m²-180g/m² an. Je mehr man sich

„Wir empfehlen ein Papiergewicht zwischen 80 und 120g/m².“

allerdings den Grenzen dieser Angaben nähert, desto größer wird die Papierstaugefahr. Wir empfehlen

daher das eigene Papier zwischen 80 und 120 g/m² zu halten. Während zu schweres Papier in der Regel gar nicht erst

gezogen wird, kann zu leichtes Papier nicht nur zu einer Stauung führen, sondern wird im schlimmsten sogar Fall beschädigt.

Papierdicke:

Die Dicke von Papier wird in der Regel nicht angegeben und findet sich auch nicht in den Spezifikationen der

meisten ADFs. Dies ist auch in der Regel nicht nötig, da die Papierdicke eine

Funktion aus Papierdicke und – Gewicht ist. Es gilt also allgemein, je schwerer mein Papier, desto dicker. Allerdings kann sich die Dichte von Papier, je nach verwendeter Sorte

erheblich unterscheiden und liegt bei einem Wert von 700 – 1200 kg/m³.

Zum Vergleich, normales

„Wir empfehlen eine Papierdicke zwischen 100 – 120µm.“

Schreibpapier hat eine Dichte von 800kg/m³. Das bedeutet, dass unter Umständen zwei Blatt

Papier mit gleichem Gewicht sich in ihrer Dicke erheblich unterscheiden können. Je geringer die Dichte desto dicker das Papier. Zum doppelseitigen Scannen muss das entsprechende

Dokument bei einigen ADFs auf kleinstem Raum gewendet werden.

Hier ist es zu dickes Papier, welches zum Fehler führt.

Papierformat:

Die Angaben der meisten Hersteller sind hier einheitlich. A4-Format in der Höhe. Grundsätzlich wird niemand eine Postkarte per Stapelverarbeitung einscannen. Es gibt jedoch Schreiben, die nur minimal von der DIN-Norm abweichen. Besonders beliebt ist zum Beispiel so genanntes Endlospapier für

Rechnungen. Im alltäglichen Betrieb bleiben die minimalen

Abweichungen von dem einzelnen Mitarbeiter unbemerkt und die Schreiben werden standartgemäß in die Stapelverarbeitung gegeben. Die Konsequenz, das Dokument passt nicht in den Einzugsschlitz des ADF und wird je nach Qualität unterschiedlich stark geknautscht oder sogar beschädigt, es kommt zum

Papierstau. Doch auch wenn es erfolgreich eingescannt wird, kann es immer noch zu Problemen in der weiteren Erfassung durch die verwendete Software kommen. Das Schreiben könnte den innerhalb der Software vordefinierten

Toleranzbereich für A4 überschreiten

und würde in diesem Fall eine Fehlermeldung provozieren. Handelsübliche DMS-Software arbeitet

standartmäßig mit eng gewählten Toleranzgrenzen um die korrekte Funktionalität einiger Features zu gewährleisten. Das Format muss dabei zusätzlich nicht nur in der Höhe, sondern auch in der Breite stimmen. Aber auch einige Scanner haben Schwierigkeiten, wenn Papiere zu sehr von normierten Formaten abweichen.

„Ihr Papier sollte die DIN Norm für A4 in der Höhe einhalten.“

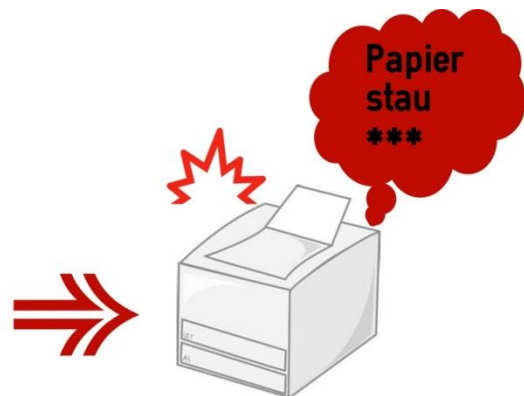
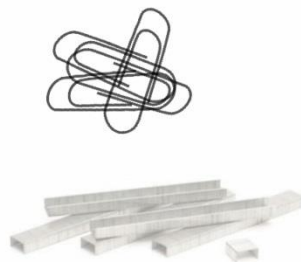
Papieroberfläche:

Papiereigenschaften wie Lichtdurchlässigkeit und Flexibilität werden in diesem Paper nicht besprochen. Beide Eigenschaften sind eine Folge aus Papiergewicht und Dicke. Bei Beachtung der vorhergehenden Ratschläge, können Fehler auf Basis dieser Eigenschaften ausgeschlossen werden. Sollte man allerdings ein anderes Medium als Papier verwenden, zum Beispiel Folie, können wiederum massive Fehler auftreten. Es ist jedoch unüblich, Folien in die Stapelverarbeitung zu geben. Sollte dies dennoch geschehen, wird die Schuld kaum beim Erzeuger des jeweiligen Schreibens gesucht werden. Zwar sind Probleme im

„Verwenden Sie raues und nicht glattes Papier.“

Zusammenhang mit der Papieroberfläche eher untypisch, dafür treten sie auf, ohne dass man vorher hätte gegensteuern können. Im normalen Bürobetrieb bleibt zum Befühlen der Dokumente keine Zeit und Schreiben mit falscher Oberfläche landen daher unweigerlich in der Stapelverarbeitung. Besonders glattes Papier wird schlecht erfasst. Die Blatteinzugsrollen drehen unter Umständen durch und es kommt zum Papierstau. Auch wenn glattes Papier ein Merkmal für besondere Qualität ist, kann es ungewollte Komplikationen verursachen und der Einsatz sollte gut überlegt sein. Raues Papier ist zumindest für Maschinenhände feiner.

Befestigungsmechanismen



Einer der häufigsten Fehler hat nicht die Qualität des verwendeten Papiers zur Ursache, sondern liegt ganz trivial in der Verwendung von Befestigungsmechanismen und zwar ganz allgemein.

Egal ob Eselsohren, Heftklammern oder jedwede Tacker, bei der Stapelverarbeitung kommt es

unweigerlich zum Stau. Das Problem ist so verbreitet, dass die meisten Geräte über einen Warnaufdruck verfügen, Heft- oder Büroklammern vor der Stapelverarbeitung zu entfernen. Im besten Fall müssen also verwendete Befestigungsmechanismen vor der automatischen Verarbeitung manuell und mühsam entfernt werden. Je nach verwendeter Technik kann das unter Umständen sehr nervenaufreibend sein. Im schlimmsten Fall werden Klammern übersehen und beschädigen das ADF-Modul. Selbst Eselsohren, welche vergleichsweise einfach zu entfernen sind, stellen in der Folge eine Schwäche des Papiers dar und führen

„Gedacht als
Organisationshilfen sind
Befestigungsmechanismen
eher organisatorische
Hürden.“

zu einem Verarbeitungsrisiko. In keinem Fall führen Befestigungen zum gewünschten Ergebnis. Was als organisatorische Arbeitserleichterung für den Partner gedacht war, ist in der

Mehrheit Arbeiterschwernis. Dabei stellt bereits der Briefumschlag das optimale Anheftungstool dar. Nach dem Öffnen, können die Dokumente

problemlos eingescannt und digital geheftet werden. Sollte es dennoch nötig sein, die Dokumente für eine analoge Archivierung zusammen zu tackern, gilt, das Setzen von Heftklammern ist schneller als deren Entfernung. Wir empfehlen Schreiben grundsätzlich ungeheftet zu verschicken.


Digitale Texterkennung – OCR

Aber selbst wenn mein Dokument erfolgreich eingescannt wurde, kann es nachfolgend zu Problemen bei der zumeist automatisierten digitalen Verarbeitung kommen. Hierbei gilt: „Nicht alles was der Mensch lesen kann, kann auch das Programm lesen.“ Dabei sollte man auf bestimmte Schlüsselbegriffe im Dokument besonderen Wert legen. Um später problemlos wiedergefunden zu werden, speichern moderne DMS-Softwares eingehende Dokumente gleich unter mehreren Schlagwörtern ab. Diese Meta-Daten werden im besten Fall automatisch dem Dokument entnommen. Egal wie die eingesetzten Verfahren dabei aussehen, sind sie immer OCR gestützt. Doch gleich wie gut die OCR, jede macht Fehler. Ein Umstand, der auch den Entwicklern von DMS-Software bekannt ist, weswegen alle nennenswerten Programme auf diesem Gebiet Überprüfungsmöglichkeiten enthalten. Solche Sicherheitsfunktionen haben jedoch ihren Preis, welchen man mit Zeit bezahlt. Möglichkeiten zur Prüfung sind somit kurz gehalten und Teil eines allgemeinen Workflows und das Risiko den Fehler zu übersehen daher vorhanden.

Zeitersparnis ist die Haupttriebfeder hinter einer DMS – Software. Die Quantität der umgesetzten Dokumente ist der Bereich innerhalb dessen sich diese Ersparnis einstellt. Was die verwendete OCR betrifft, entsteht unweigerlich ein Konflikt zwischen der angestrebten Quantität der umgesetzten Dokumente pro Minute und einer möglichen Qualität der Ergebnisse. Viele Optimierungsvorgänge von OCR – Stand – Alone – Programmen, wie etwa die automatische Kontrastierung, sind zu rechenlastig und in der Regel abgestellt. Die Fehler, die durch solche Optimierungen verhindert werden, betreffen nur einzelne Dokumente und wären bei der Masseverarbeitung unnötig angewandt. Ein reines OCR- Programm vermag 10 Seiten pro Minute zu bewältigen. Für moderne DMS ist das noch zu langsam. Umso wichtiger ist es, bereits bei der Erstellung von Schreiben die optische Zeichenerkennung zu berücksichtigen. Die folgende Übersicht fasst alles zusammen,

was man beachten muss, um die eigenen Schreiben auf OCR hin zu optimieren:

OCR Optimierung





Header

Logos kann ich nicht lesen



- enthält alle relevanten Daten wie zum Beispiel:
- befindet sich auf Seite 1
- besitzt ein statisches Layout


Kundennummer
 Datum Firmennamen
 Rechnungsnummer
 Adressat Vertragsnummer
 Projektbezeichnung
 Firmenanschrift Ansprechperson
 Gesamtseitenzahl Dokumenttyp ...

Abstände beachten !!!


Fließtext

- den Kontrast beachten nicht nur schwarz - weiß
- die Schriftgröße sollte einheitlich sein, weder zu  noch zu 
- Schriftarten verwenden, die Druckschrift möglichst ähneln
- Sprache einer Seite einheitlich halten



Не смешивать языки !

Die russische Übersetzung ist auf Seite 7!



Wo ist die Seitenzahl ???

Header

Das wohl wichtigste Feature moderner DMS – Software ist die automatische Verschlagwortung und die Schlagwortsuche im Archiv. Dabei werden relevante Informationen aus dem Text ausgelesen und in Form von Metadateien an das archivierte Dokument angehängt. Unter relevanten Informationen versteht man alles, wonach später wahrscheinlich gesucht werden könnte. Ist ein Dokument erst ins System eingelernt worden, werden alle späteren Dokumente desselben Typs selbständig erkannt und verschlagwortet. Dabei dienen immer gleiche Informationen, wie z.B. Dokumenttyp und Absender dem Programm als erster Anhaltspunkt, um das Schreiben richtig zuzuordnen.

„Das Gesicht eines Dokumentes, ist das woran man es wiedererkennt.“

Über Variablen wie Fixinformationen, die Positionierung dieser innerhalb des Dokumentes und Volltextübereinstimmung wird ein statistischer Wert ermittelt, auf dessen Basis veränderbare, relevante Informationen ausgelesen werden.

Auch wenn sich die einzelnen Verfahren von Programm zu Programm unterscheiden, spielen diese Variablen stets eine Rolle. Zum Auslesen wiederum ist im Programm die Position der

Zielinformation mittels Vektoren hinterlegt. Schlüsselbegriffe wie Datum, können zusätzlich dabei helfen, Vektoren an das vorliegende Dokument anzupassen.

Konkret bedeutet das, dass bei der Erstellung eines Dokumentes auf folgendes zu achten ist:

Meine Dokumente sollten über ein statisches Layout verfügen, in dem alle relevanten Informationen enthalten sind.

Tauchen relevante Informationen im Fließtext auf, kann sich ihre Position

von Dokument zu Dokument unter Umständen deutlich unterscheiden. Im

besten Fall werden solche Informationen falsch ausgelesen und im schlimmsten reicht die stochastische Übereinstimmung für die Software nicht aus, um das Dokument

richtig zuzuordnen und es muss manuell verschlagwortet werden. Das Springen relevanter Information im Text bedeutet aber auch, dass auch ein Mensch nach diesen suchen muss.

Alle relevanten Informationen sollten sich bereits auf der ersten Seite finden lassen.

Auch wenn viele DMS oder ECM – Systeme die Möglichkeit einer Weiterbearbeitung bieten, die seitenübergreifend funktioniert, spielt dies bei der Ersterkennung keine Rolle und wäre für den praktischen Betrieb auch dysfunktional. Schließlich dient

die Titelseite auch in analogen Systemen primär der sauberen Archivierung. Relevante Informationen, die sich nicht auf der ersten Seite befinden, werden daher in der automatischen Verschlagwortung nicht länger berücksichtigt.

Relevante Informationen sollten über genügend Abstand zu anderem Text verfügen.

Diese unbestimmte Angabe bezieht sich auf das leichte Spiel, das papiergebundene Schreiben während des Scanvorganges haben. Eventuelle minimale Verschiebungen können dazu führen, dass benachbarte Zeichen ins vordefinierte Feld für die relevanten Informationen geraten und zu Auslesefehlern führen. Je

größzügiger dieses Feld gewählt werden kann, desto kleiner das Risiko falsch zu verschlagworten. In der Regel reichen vier bis fünf Informationen aus, um ein Dokument schnell finden zu können. Durch den Abstand bestimmen Sie mit, welche Informationen Ihr Gegenüber einlernen wird.

Nachtrag

Logos werden in der Regel schlecht erkannt und die in ihnen enthaltenen Informationen sollten dementsprechend erneut innerhalb des Dokumentes auftauchen.

Fließtext

Auch wenn die Aussagen zum Fließtext primär auf die Volltexterkennung und somit auf die Volltextsuche zielen, sind diese genauso auf relevante Informationen anzuwenden. Es ist zwar trivial, doch wird dennoch oft vergessen: „Nicht alles was der Mensch lesen kann, kann auch die Maschine auslesen.“

Schriftgröße

Während die Schriftgröße bei den gängigen Textdateien keine Rolle für die OCR spielt, kommt es bei Bilddateien schon ab einer Schriftgröße von 8 und kleiner zu Fehlern. Dies liegt vor allem an der Auflösung, mit der die Dokumente eingescannt werden. Während diverse Stand - Alone - OCR - Tools die Möglichkeit der Nachbesserung bieten, sind solche Funktionen bei DMS - Softwares eher

abgestellt. Zu große Schrift produziert aus den gleichen Gründen einen ähnlichen Fehler. Obwohl der Buchstabe für das menschliche Auge gut lesbar erscheint, sieht das Programm auch die Auflösungsfehler und wird einem einzelnen Buchstaben unter Umständen mehrere Zeichen zuordnen. Wir empfehlen wichtige Informationen max. die Schrittgröße 16 zuzuordnen.

„Wir empfehlen eine Schriftgröße von 8 -16.“

Kontrast

Ein weiterer Faktor bei der Texterkennung ist die Kontrastierung mit dem Hintergrund des Dokumentes.

Wiederrum sind Möglichkeiten der nachträglichen Aufhellung bei DMS – Programmen nicht möglich. Zwar stellt Schwarz – Weiß den optimalen Kontrast

dar, dennoch ist man bei der Gestaltung seiner Schreiben nicht auf diesen festgelegt. Grundsätzlich ist jede Farbkombination denkbar, die einen deutlichen Kontrast produziert. Ein farblicher Hintergrund birgt jedoch die Gefahr, sollte das Dokument nur in

Schwarz – Weiß eingescannt werden, dass sich der Kontrast in der Folge

abschwächt. Diese Gefahr wird allerdings relativiert, wenn man bedenkt, dass das Scannen in Farbe grundsätzlich die Qualität der Weiterverarbeitung erhöht. Zum Beispiel hat eine Schwarz-Weiß-Bilddatei mit ihren, in der Regel, 8 – Bit nicht die

„Der farbliche Hintergrund der Seite sollte mit der Farbe der verwendeten Zeichen einen deutlichen Kontrast bilden. Auf Schwarz-Weiß-Scannen braucht keine Rücksicht genommen zu werden.“

nötigen Farbdimensionen um mit einem digitalen Farbstempel versehen zu werden. Praktisch scannen alle in Farbe, spätestens nach dem ersten Support- Anruf beim jeweiligen DMS-Anbieter.

Schriftart

Bei der Wahl der eingesetzten Schriftart, sollte man sich auf die Standardschriftarten beschränken. Während Stand – Alone Produkte über verschiedene Datenbanken verfügen und sogar die Möglichkeit bieten neue

Schriftarten in verschiedenen Sprachen hinzuzufügen, entfallen solche Dienste bei den verwendeten ORC-Tools von DMS- Software. Aber auch Schriftarten die Schnörkel oder Ähnliches enthalten, könnten unter Umständen

falsch ausgelesen werden. Grund hierfür ist die Auflösung beim Scannen in Kombination mit der Überschneidung der Formdefinitionen solcher Zeichen. Ebenso wird vor allem kursiv aber auch fett und unterstrichen formatierter Text schlechter erkannt. Wenn man etwas hervorheben möchte, sollte man lieber die nächsthöhere Buchstabengröße verwenden. Die besten Ergebnisse liefern Zeichen, die Druckbuchstaben

„Die verwendeten Zeichen sollten Druckbuchstaben nach Möglichkeit ähneln, weder kursiv, fett oder unterstrichen sein.“

stark ähneln. Da Druckbuchstaben sich nur aus den beiden Elementen Halbkreis und Gerade zusammensetzen, besitzen sie in einem linguistisch – ökonomischen Sinne die größte Informationsdichte. Während für die Erkennung eines Zeichens irrelevante Schnörkel zu Fehlbestimmungen führen können, fehlen solche überflüssigen Gestaltungselemente bei Druckbuchstaben völlig.

Sprachmischung

Neben den Schriftarten ist auch ein Wörterbuch in mehreren Sprachen im Hintergrund einer OCR hinterlegt. Nach dem Auslesen der einzelnen Zeichen einer Zeichengruppe, werden diese anhand statistischer Werte einem Wort zugeordnet. Es kann also sein, dass selbst wenn die OCR die Zeichen * H u n b * ausliest, im endgültigen Ergebnis die Gruppe dennoch richtig

„Die Sprache einer Seite sollte einheitlich sein.“

unter Hund zusammengefasst ist. Dieses Feature bedeutet eine enorme Verbesserung in den Ausgabeergebnissen einer OCR, verursacht jedoch bei Sprachwechseln Fehler. Deswegen legt sich die OCR bereits im Vorfeld einer Seitenanalyse auf eine Sprache fest. Einzelne Wörter einer anderen Sprache werden in der Regel falsch erkannt.

Sprachvermischungen auf einer Seite sollte man daher meiden.

Nachtrag

Bei digital versendeten Dokumenten sollte stets die Datei auch im pdf – Format korrektur gelesen werden. Zum Zwecke der Archivierung werden alle Formate grundsätzlich automatisch von einer DMS – Software im pdf – Format gespeichert. Bei der Umwandlung können ungewollte Fehler auftreten, die sich negativ auf eine automatische Texterkennung auswirken.