EIN NEUER BLICK AUF EINE HOCHENTWICKELTE KOMPRIMIERUNG GESCANNTER DOKUMENTE

Optimierte PDFs bieten Vorteile, die sowohl der heutigen als auch der zukünftigen IT-Landschaft gerecht werden

WHITEPAPER

INHALTSVERZEICHNIS

Ein neuer Blick auf eine hochentwickelte Komprimierung gescannter Dokumente 3

Spezielle Vorteile der hochentwickelten Komprimierung 4

Funktionsweise 7

Eine ergänzende Technologie 8

Schlussfolgerung 8

Über den Autor 9

Fußnoten 9

EIN NEUER BLICK AUF EINE HOCHENTWICKELTE KOMPRIMIERUNG GESCANNTER DOKUMENTE

Wenn es um digitale Dateien geht, war kleiner schon immer besser. Kleinere Dateien benötigen weniger Speicherplatz und weniger Zeit zum Herunterladen, verbrauchen weniger Netzwerkbandbreite und lassen sich leichter per E-Mail versenden und austauschen. In vielen Fällen jedoch, vor allem bei gescannten Dokumenten, arbeiten Sie wahrscheinlich mit Dateien, die 10 bis 50 Mal größer sind, als sie sein müssten.

Wenn Sie keine hochentwickelte Komprimierungstechnologie in Ihre Dokumentenerfassung integrieren, haben Sie überflüssige Kosten für Speicherplatz, verbrauchen mehr Netzwerkbandbreite, füllen Mailboxen schneller und verlangen von anderen Menschen, länger als nötig auf den Download von Dateien zu warten. Der PDF Compressor von Foxit löst nicht nur diese Probleme, sondern behandelt auch völlig neue Aspekte in Bezug auf die Dateigröße, die mit der zunehmenden Entwicklung hin zur Mobiltechnologie entstehen.

In der Vergangenheit haben die Anwender Dokumente von Websites auf ihren PC heruntergeladen oder als E-Mail-Anhang versendet und empfangen. Zum Übertragen besonders großer Dateien wurde auch FTP genutzt. In den letzten fünf Jahren ist es jedoch zu einer explosionsartigen Zunahme von Smartphones, Tablets und Cloud-Storage gekommen. Diese neuen Kommunikationswege erweitern die Art und Weise, wie wir Dateien nutzen, teilen und speichern.

Anstatt Dateien nur über kabelgebundene Internetverbindungen herunterzuladen, öffnen wir jetzt Dokumente auf mobilen Computern über Mobilfunknetze oder laden sie auf Cloud-Server für die gemeinsame Nutzung und Zusammenarbeit hoch.

Mit dieser Entwicklung hin zum Zeitalter der portablen Dokumente wurde die Größe der Dateien wichtiger denn je. Niemand wartet gerne übermäßig lange auf einen Download, aber die oft wechselnden Geschwindigkeiten von Mobilfunkverbindungen haben die Probleme bei der Arbeit mit großen Dateien noch verschärft. Zusätzlich sorgen mobile Datentarife mit begrenztem Kontingent pro Monat und hohen Kosten für zusätzliches Datenvolumen sowie Anbieter von Cloud-Speicherplatz, die nach GB abrechnen, dafür, dass die Minimierung der Dateigrößen wichtiger als je zuvor ist.

Im Jahr 2012 kündigte der schwedische Telekommunikationsriese Ericsson an, dass er an einer neuen Komprimierungstechnologie arbeitet, die speziell darauf ausgerichtet ist, „den Bandbreitenbedarf für die Videoauslieferung um mehr als 50 Prozent zu reduzieren."[1] Für Unternehmen, die Dokumenten-Imaging einsetzen, gibt es die gute Nachricht, dass es bereits jetzt eine Technologie auf dem Markt gibt, die gescannte Dokumente um weit mehr als 50 Prozent komprimieren kann – und dies mit einer Geschwindigkeit von mehreren Tausend Seiten pro Stunde.

Foxits PDF Compressor ist eine ausgereifte und bewährte Anwendung, die von Tausenden von Unternehmen weltweit eingesetzt wird, um nicht nur die Größe von PDF-Dateien zu reduzieren, sondern auch, um hochpräzise Bilder zu erstellen, die eine Volltextsuche ermöglichen. Der PDF Compressor arbeitet mit verschiedenen Technologien, um Kompressionsraten von 10 zu 1 bei einfacheren, textlastigen Schwarzweißbildern, mehr als 100 zu 1 bei grafiklastigen Farbdateien und OCR-Genauigkeitsraten von über 99 % (gemessen an der Zeichenanzahl) zu erreichen. Zu diesen Technologien gehören eine hochentwickelte Bildverarbeitung, Graustufenschwellwerte und das MRC-Verfahren (Mixed Raster Content).

Der in den Erfassungsvorgang integrierbare PDF Compressor bietet vielfältige Vorteile, darunter potenziell drastische Kostensenkungen für die Speicherung der Dokumentenbilder – eine Einsparung, die allein schon die Kosten für die Software rechtfertigt. Der PDF Compressor ergänzt bestehende Erfassungs- und Dokumentenmanagementsysteme, weil das Produkt die Funktionalität beider Anwendungen erweitert. Zusätzlich macht es sie kompatibler mit der aufkommenden Welt der portablen Dokumente.

SPEZIELLE VORTEILE DER HOCHENTWICKELTEN KOMPRIMIERUNG

Sehen wir uns genauer an, welche Vorteile PDF Compressor für ein Dokumenten-Imaging-System bieten kann:

1. Geringere Speicherungskosten:

Vor-Ort-Speicher

Cloud-Speicher

In den letzten Jahren haben die sinkenden Kosten für Festplattenspeicher dazu geführt, dass die wahren Speicherungskosten oft unterschätzt werden. Auch wenn es stimmt, dass der Preis für Speicherplatz auf Festplatten, die in großen Stückzahlen gekauft werden, nur noch wenige Cent pro GB beträgt, entspricht dies bei weitem nicht den tatsächlichen Kosten für die Speicherung. Wenn Sie die Kosten für Server, Backups, Systemwartung, Personal, Stromversorgung und Räumlichkeiten berücksichtigen, liegen die tatsächlichen Kosten für Vor-Ort-Speicher wahrscheinlich eher bei 10 Dollar pro GB und Jahr[2], es gibt aber auch Berichte von IT-Abteilungen, die sogar mit 30 Dollar pro GB und Monat kalkulieren[3]. Auch wenn Daten-Hosting-Giganten wie Amazon und Google mit 0,10 Dollar pro GB und Monat werben, umfasst dies nur die simple Speicherung, aber kein Dokumentenmanagementsystem. Jedes Unternehmen, das geschäftskritische Dateien speichert, ist in der Regel geneigt, ein gehostetes Dokumentenmanagementsystem zu verwenden. Abhängig von Variablen wie Nutzungsgebühren, Mengen und Features können die monatlichen Kosten für gehostetes Dokumentenmanagement zwischen 2 und 50 Dollar pro GB liegen.

Der PDF Compressor kann Dateigrößen und Speicherungskosten erheblich reduzieren. Angenommen, Sie speichern TIFF-Dateien im Umfang von 1 TB. Bei geschätzten Kosten von 10 Dollar pro GB und Jahr für den Vor-Ort-Speicher (die wahrscheinlich am unteren Ende liegen), würden Sie ungefähr 10.000 Dollar pro Jahr für die Speicherung dieser Dokumente bezahlen. Basierend auf Tests, die Foxit mit Beispieldokumenten durchgeführt hat, kann durch den Einsatz des PDF Compressors und die Konvertierung dieser TIFFs in PDF-Dateien, in denen nach Text gesucht werden kann, deren Größe um ca. 90 % reduziert werden.

Aus 1 TB Dokumentenspeicher werden 100 GB, wodurch die Kosten für Vor-Ort-Speicher von 10.000 Dollar pro Jahr auf 1.000 Dollar sinken. Zugegeben, dies sind „potenzielle“ Einsparungen und sie setzten voraus, dass die freigewordenen Ressourcen im Speichersystem vollständig umverteilt werden können (beispielsweise zum Speichern weiterer Dokumente). Aber auch wenn nur die Hälfte der potenziellen Einsparungen tatsächlich realisiert werden können, sieht man deutlich, wie der PDF Compressor sich selbst bezahlt machen kann.

In der aufstrebenden Welt des Cloud-Storage können die Einsparungen noch ausgeprägter sein.

Angenommen, der obige Anwender zahlt 10 Dollar pro GB und Monat für cloudbasiertes Dokumentenmanagement. Dies sind 120 Dollar pro Jahr und GB bzw. 60.000 Dollar für 5 TB Dokumentenspeicherplatz. (Da Cloud-Storage in der Regel ein Backup beinhaltet, wird nicht so viel Rohdatenspeicher benötigt, um das gleiche Volumen an Dokumenten wie bei Vor-Ort-Speicher zu bewältigen.) Durch den Einsatz des PDF Compressors (mit einer Komprimierungsrate von 90 % für schwarzweiße TIFF-Dateien) kann der Speicherbedarf für das Cloud-Dokumentenmanagement von 500 GB auf 50 GB reduziert werden, was einer potenziellen Senkung der Cloud-Storage-Kosten von 60.000 auf 6.000 Dollar entspricht.

Werden Sie jetzt durch die Komprimierung von 0,5 TB an Dokumenten wirklich jährlich 54.000 Dollar Einsparungen beim Cloud-Storage erreichen? Wahrscheinlich nicht, da oft auch Benutzerlizenzen in die Kosten für cloudbasierte Speicher einfließen. Außerdem werden Sie bei den meisten Preismodellen der Cloud-Anbieter für geringere Speichervolumina deutlich mehr pro GB bezahlen als für ein halbes Terabyte oder mehr – wahrscheinlich mindestens doppelt so viel. Dennoch können die mit dem PDF Compressor erreichten Einsparungen an Cloud-Speicherplatz erheblich sein und die Kosten für die Software mehr als decken. Dies könnte auch der entscheidende Faktor sein, der Ihr Unternehmen in die Lage versetzt, eine bestehende Dokumentenmanagement-Anwendung in eine modernere, einfacher zu wartende Cloud-Umgebung zu überführen.

Diese Einsparungen erhöhen oder verringern sich natürlich je nach Umfang und Art der Dokumente. Bei Farbdokumenten beispielsweise, bei denen die Komprimierungsrate durchschnittlich 98 % betragen kann, können die Einsparungen sogar noch höher sein.

2. Verbesserte Möglichkeiten der Zusammenarbeit:

Das Versenden gescannter Dokumente per E-Mail ist eine gängige und bequeme Möglichkeit, Papierdokumente auszutauschen. Nach Angaben der Radicati-Gruppe wurden im Jahr 2012 täglich 89 Milliarden geschäftliche E-Mails verschickt. Laut Osterman Research enthalten 29 % der geschäftlichen E-Mails Anhänge, was bedeutet, dass im Jahr 2012 täglich rund 25 Milliarden E-Mail-Anhänge verschickt wurden.

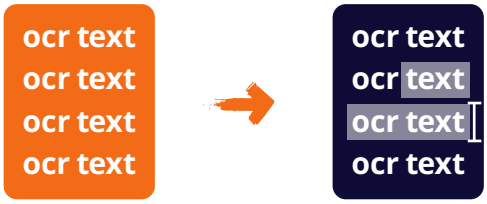
Der PDF Compressor ist beim Versenden von gescannten Dokumenten doppelt nützlich. Erstens wird sichergestellt, dass die gescannten Bilder in einem einheitlichen Format – PDF – vorliegen, das die meisten Empfänger ohne speziellen Viewer öffnen können. Und zweitens wird gewährleistet, dass Sie die geringstmögliche Dateigröße versenden und empfangen. Dadurch werden nicht nur Netzwerkbandbreite, E-Mail-Server und Desktop-E-Mail-Anwendungen entlastet, sondern auch die Nutzer profitieren von höheren Geschwindigkeiten beim Download und Öffnen von Dateien.

Der Aspekt der Bandbreite wird beim Dateizugriff mithilfe von mobilen Endgeräten noch wichtiger. Viele Datentarife begrenzen das monatliche Datenvolumen der Kunden. Der Download eines 6 MB großen, unkomprimierten JPEG-Bilds eines Dokuments verbraucht 17 Mal mehr Bandbreite als der Download einer 357-KB-Version derselben Datei, die durch den PDF Compressor gelaufen ist.

In einer Nielson-Umfrage wurde festgestellt, dass sich das verbrauchte Datenvolumen eines durchschnittlichen Datentarif-Kunden im Jahr 2012 gegenüber 2011 verdoppelt hat[4], wobei sich weiteres Wachstum abzeichnet. Bei mehreren Geräten, die unter einem gemeinsamen Unternehmensdatentarif laufen, verschlimmern sich die Bandbreitenprobleme noch.

Im Endeffekt will niemand unnötig Bandbreite verbrauchen, die später für andere Aufgaben benötigt wird. Darüber hinaus können Überschreitungen des Datentarifs nicht nur teuer, sondern auch hinderlich werden, wenn der Anbieter die Übertragungsrate verlangsamt, nachdem die Nutzer ihr Kontingent ausgeschöpft haben. Die Nutzung von PDF Compressor ermöglicht Ihrem Unternehmen, Dateien zu erstellen, die 10 % oder weniger der Bandbreite von unkomprimierten Dateien benötigen. Dies gilt für Servicehandbücher, Formulare, Berichte, Rechts- und Finanzdokumente – eigentlich für jede Datei, die ein Mitarbeiter per E-Mail versenden oder über ein Handy oder Tablet nutzen möchte.

3. Vollständig durchsuchbare PDF-Dateien:



Der PDF Compressor enthält modernste OCR-Technologie für die Erstellung von im Volltext durchsuchbaren PDFs. Die OCR-Ergebnisse sind in der PDF-Datei enthalten, wodurch sie mit der Datei portabel sind. Im Gegensatz dazu müssen als TIFFs und JPEGs gespeicherte Dokumentbilder mit separaten Datendateien verknüpft werden, die die OCR-Ergebnisse enthalten.

Der PDF Compressor wendet eine proprietäre Bildverarbeitung an, um OCR-Ergebnisse zu erzielen, die laut Tests von Foxit 15 % genauer (auf Zeichenebene) als die Ergebnisse einer führenden OCR-Engine sind. Eine hochpräzise Wiedergabe des Textes in Verbindung mit einem Dokumentbild ist besonders wichtig für Anwendungen in Bereichen wie dem Rechtswesen, wo oft unterschiedliche Dokumenttypen ohne vordefiniertes Indizierungssystem verwendet werden. Die Volltextsuche kann bei der Archivierung wichtig sein, da Sie oft nicht wissen, welche Arten von Recherchen in den nächsten Jahren durchgeführt werden sollen – speziell für Angelegenheiten rund um e-Discovery.

Für die Archivierung verfügt der PDF Compressor zudem über die Möglichkeit, PDF/A-Dateien auszugeben. Dabei handelt es sich um einen ISO-Standard zur langfristigen Sicherstellung der Zugänglichkeit. PDF/A ist in vielen europäischen Ländern als Archivierungsstandard anerkannt und findet auch Einzug in Nordamerika, insbesondere in regulierten Branchen.

FUNKTIONSWEISE

PDF Compressor verfügt über mehrere anpassbare Einstellungen. Sie lassen sich in fünf Hauptkategorien einteilen:

Bildverarbeitung
PDF Compressor verfügt über eine Reihe von Optionen zur Bereinigung von Bildern. Das Produkt bietet auch Schwellenwerte für Graustufen- und Farbbilder, wodurch der Kontrast zwischen Vordergrund und Hintergrund erhöht wird, um optimale bitonale Dateien zu erzeugen.

MRC-Segmentierung
Beim MRC-Verfahren wird eine Datei vor der Komprimierung in Ebenen aufgeteilt. Mit MRC kann die höchstmögliche Komprimierung bei Farbscanns erreicht werden. MRC trennt die Dateien in Text- und Grafikebenen und komprimiert anschließend die einzelnen Ebenen unterschiedlich.

Komprimierung
Der PDF Compressor enthält Optionen für vier verschiedene Komprimierungsmethoden. Für Farb- oder Graustufendokumente und/oder Grafikelemente von MRC-Dokumenten stehen die JPEG- und die JPEG-2000-Komprimierung zur Verfügung. JBIG2 und Group 4 sind für Schwarzweiß-Dokumente und/oder Textelemente von MRC-Dokumenten verfügbar.

OCR
Auf der einfachsten Ebene können Sie OCR ein- und ausschalten. OCR-Daten erhöhen die Dateigröße nur unwesentlich (ca. 5 %), verlangsamen aber den Konvertierungsprozess (Details siehe unten). Bis zu 117 verschiedene Sprachen können bei der Ausführung des OCR-Prozesses berücksichtigt werden. Der PDF Compressor bietet vier OCR-Modi. Um bei der Entscheidung zu helfen, welcher Modus für Ihre Anwendung der richtige ist, kann die Software Berichte über das Konfidenzlevel der Genauigkeit für eine Reihe von Dokumenten erstellen, die in einem bestimmten Modus verarbeitet werden. Weiterhin kann Der PDF Compressor Daten darüber liefern, wie lange die Erzeugung eines bestimmten Typs einer durchsuchbaren PDF-Datei dauert.

Geschwindigkeit
Bei Tests von Foxit konnte der PDF Compressor ohne OCR rund 40.000 Seiten pro Stunde und Prozessorkern ausgeben, mit OCR waren es ca. 3.000 Seiten pro Stunde und Prozessorkern. Der PDF Compressor ist darauf ausgelegt, die Vorteile einer Quad-Core-CPU vollständig auszuschöpfen. Jeder Kern kann eigenständig Seiten verarbeiten. Daher vervierfachen sich die angegebenen Geschwindigkeiten bei einem Quad-Core-Prozessor. Bestimmte OCR-Einstellungen wirken sich ebenfalls auf die Ausgaberaten aus.

Der Schnelleinrichtungsassistent des PDF Compressors stellt eine Reihe von Fragen und schlägt auf Basis Ihrer Antworten die optimalen Einstellungen vor. Es werden Beispielausgabedateien erstellt, zusammen mit Informationen darüber, wie lange es

dauerte, die einzelnen Beispieldateien zu generieren. Wenn Sie mit einem dieser Beispiele zufrieden sind, können die Einstellungen als Profil gespeichert werden. Falls nicht, stellt der Assistent zusätzliche Fragen und generiert einen neuen Satz von Beispielen.

EINE ERGÄNZENDE TECHNOLOGIE

Der PDF Compressor ist eine hervorragende Ergänzung für jede Dokumenten-Imaging-Implementierung. Das Produkt ist zwischen der Erfassung und dem Dokumentenmanagement angesiedelt und trägt dazu bei, die Leistung beider Anwendungen zu optimieren. Es hilft der Erfassungssoftware dabei, ein höhere Qualität auszugeben, und sorgt für einen effizienteren Betrieb der Dokumentenmanagementsoftware, da sie weniger Speicherplatz benötigt und kleinere Dateien zum Download anbietet.

Der PDF Compressor kann über vordefinierte Schnittstellen direkt in führende Erfassungsanwendungen wie Kofax Capture und EMC Captiva integriert werden. Auf der Befehlszeilenebene ist eine Integration des PDF Compressorsin andere Erfassungs- und Dokumentenmanagementanwendungen möglich. Das Produkt kann auch als eigenständiger Prozess konfiguriert werden, der automatisch Dateien aus einem überwachten Ordner einliest, konvertiert und in das Dateisystem ausgibt.

Der PDF Compressor eignet sich sowohl für neue als auch für bestehende Anwendungen im Bereich Dokumenten-Imaging. Für neue Anwender einer Dokumentenmanagementtechnologie bietet es die Gelegenheit, ein System zu erhalten, das im Hinblick auf den Speicherplatz als auch auf den Download für mobile Endgeräte optimiert ist. Für Anwender älterer Systeme bietet es eine relativ kostengünstige Möglichkeit, ein System zu modernisieren und aufzufrischen, ohne in ein neues Dokumenten-Imaging-System investieren zu müssen.

Auch zum Umstieg von Altanwendungen auf Cloud-Umgebungen ist der PDF Compressor ein hervorragendes Werkzeug. Wenn Sie bisher gezögert haben, von einem alten Dokumentenmanagementsystem (für das der Hersteller vielleicht kaum noch Support bietet) in eine modernere Cloud-Umgebung zu wechseln, weil die Kosten für die Speicherung der Altdokumente in der Cloud zu hoch sind, können diese Kosten mithilfe von PDF Compressor deutlich reduziert werden.

Durch die Implementierung von PDF Compressor können Sie auch die Vorteile der fortschrittlichsten Scannertechnologien besser nutzen. Die meisten Hochgeschwindigkeits-Dokumentenscanner sind heute in der Lage, qualitativ hochwertige Farbbilder mit der gleichen Geschwindigkeit zu erzeugen, mit der sie auch bitonale und Graustufenbilder produzieren. Viele Unternehmen nutzen diese Möglichkeit, die beste Darstellung ihrer Dokumente zu erzeugen und zu speichern, jedoch nicht, da sie Bedenken bezüglich der Speicherung der größeren Farbdateien haben. Dank der Fähigkeit von PDF Compressor, Farbscanns um mehr als 95 % zu verkleinern, sollten derartige Bedenken ausgeräumt sein, und Sie können die Vorteile Ihrer Scanner voll ausschöpfen.

SCHLUSSFOLGERUNG

Der PDF Compressor bietet eine hochentwickelte und ausgereifte Technologie, die speziell auf viele bestehende und neue

Anforderungen von Unternehmen zugeschnitten ist, die Dokumenten-Imaging-Technologie einsetzen. Mit dem PDF Compressor können Sie kleinere, nach Text durchsuchbare PDF-Dateien erstellen, die aus den folgenden Gründen von Nutzen sind:

Geringere Speicherungskosten:
Aufgrund des Wissens um ständig fallende Festplattenpreise erkennen viele Menschen nicht, wie hoch die wahren Kosten für Speicher wirklich sind. Wenn Sie die Kosten für Server, Backups, Systemwartung, Personal und Räumlichkeiten berücksichtigen, summieren sich die Speicherungskosten. Noch deutlicher werden die Einsparungen, wenn Anwender ihre Dokumente auf Cloud-Speicher verschieben, wo oft ein Entgelt pro genutztem GB anfällt. PDF Compressor ist in der Lage, die Größe von Dokumentenbilder um mehr als 90 % zu reduzieren, wodurch die Speicherungskosten potenziell um einen ähnlichen Prozentsatz sinken können.

Kleinere Datei-Downloads
Wenn kleinere Dateien erzeugt werden, erhöht dies die Arbeitsgeschwindigkeit beim Austauschen der Dateien per E-Mail. Besonders wichtig wird dies, je mehr die Anwender auf Mobilgeräte wechseln, bei denen das Datenvolumen beschränkt ist und durch große Dateien schnell aufgezehrt wird.

Präzise durchsuchbare Dateien
PDF Compressor liefert OCR-Ergebnisse, die die Branchenstandards übertreffen. Die Texte können innerhalb der PDF-Datei gebündelt werden. PDF/A-Dateien erhalten die Fähigkeit zur Textsuche über lange Zeit.

ÜBER DEN AUTOR

Ralph Gammon ist seit 2002 Redakteur und Herausgeber des Newsletters „Document Imaging Report“. Zuvor war er fünf Jahre lang Redakteur bei Jameson Publishing mit dem Schwerpunkt auf Dokumenten-Imaging und ECM-Branchen. Ralph Gammon erinnert sich noch an die Zeiten, in denen PDF als proprietäres Format angesehen wurde, das sich niemals durchsetzen würde. Neben dem Newsletter ist Gammon Verfasser des populären Blogs „Document Imaging Talk“ (http://documentimagingreport.blogspot.com/) und tweetet regelmäßig unter @DIREditor.

FUßNOTEN

1. http://www.infoworld.com/d/mobile-technology/new-video-compression-technology-willcut-bandwidth-use-in-half-says-ericsson-200747
2. File Storage Costs Less In The Cloud Than In-House, by Andrew Reichman, Forrester Research, 2011.
3. http://hardware.slashdot.org/story/10/07/29/1952235/internal-costs-per-gigabyte-whatdo-you-pay
4. http://www.phonearena.com/news/Heres-the-amount-of-data-used-by-the-average-U.S.-subscriber_id32592

Foxit
Landgrafenstr. 14
10787 Berlin, Deutschland

Vertrieb: +49 30 21 78 36 90
oder +49 30 39 40 50 20
Support: +49 30 21783691

Support Center
www.foxit.com

© Foxit Software Incorporated. Alle Rechte vorbehalten.