



Portable Document Format

INHALT

Dokumente im PDF-Format	2
Preis- und Aufwandskalkulationen auf PDF-Basis.....	2
Layout.....	2
PDF Formate als Grafik.....	3
Handhabung und Aufwand	3
Allgemeine Informationen zu PDF.....	4
Portable Document Format (PDF)	4
Verwendung und Vorteile	4
Eigenschaften von PDF-Dateien	6
Sicherheitslücken durch menschliches Versagen.....	6
Vermeidung.....	6
PDF in Betriebssystemen	7
Boxen im PDF	7
MediaBox.....	7
BleedBox.....	7
TrimBox.....	8
Art Box (auch Bounding Box).....	8
PDF in der Druckvorstufe	8
Programme zur Erstellung von PDF für die Druckvorstufe	8
Programme zum Betrachten von PDF-Dateien	9
Programme zum Erstellen von PDF-Dateien.....	9



Dokumente im PDF-Format

Immer mehr Kunden gehen dazu über, das praktische Dateiformat PDF als Ausgangsdatei für die Übersetzung zu wählen.

Was zunächst – und in diesem Rahmen des Dateiaustausches immer noch – als ein flexibles, system- und programmunabhängiges Format gilt, hat für die Übersetzungsbranche so ihre Tücken.

Zwar ist es möglich den Text aus dem PDF-Format in ein anderes – wie z.B. Word – zu bringen, doch wie so häufig, liegt der Teufel im Detail...

Preis- und Aufwandskalkulationen auf PDF-Basis

Für eine einfache Textkalkulation, d.h. die Ermittlung der Zeilen- oder Wortzahl, wird der Text exportiert und ins Word-Format gebracht. Dies kann beispielsweise im Adobe Reader über die Funktion „In Zwischenablage kopieren“ erfolgen.

Hier muss der Text dann noch überarbeitet werden, vorhandene Formatierungen müssen gelöscht und überflüssige Textelemente wie z.B. Zahlenreihen oder Absatzmarken entfernt werden, damit eine faire Preiskalkulation garantiert werden kann.

Zur weiteren Bearbeitung für die Übersetzung, insbesondere unter Verwendung von Translation Tools wie z.B. Trados oder Transit, ist aber Vorsicht geboten, da die Texte nicht in einer logischen Reihenfolge übernommen werden. Das bedeutet, je komplizierter der Text aufgebaut ist – z.B. über mehrere Spalten, viele Tabellen etc. – desto mehr Vorarbeit ist notwendig, um eine richtige und überschaubare Textreihenfolge zu erlangen. Im ungünstigsten Fall muss jeder Satz kontrolliert und evtl. an einen anderen Platz gestellt werden...

Layout

Was passiert mit dem Layout? Bei dieser Bearbeitungsmethode geht das Layout komplett verloren und muss im neuen Format nachgearbeitet werden; Abbildungen und Graphiken können nicht übernommen und müssen gesondert in das neue Dokument gebracht werden.

Oftmals bietet es sich an, den Text im Tabellenformat zu erfassen, wenn der Kunde nur die Übersetzung benötigt und das Layout extern in einer anderen Anwendung als MS-Word oder OpenOffice, wie z.B. QuarkXpress oder Pagemaker, weiterverarbeitet wird. Durch diese Methode kann sichergestellt werden, dass der Layouter die richtige Übersetzung für die jeweilige Textpassage übernimmt.



PDF Formate als Grafik

Sind aber einige Seiten aus dem vorliegenden PDF als Graphiken eingescannt worden, sieht die Sache schon anders aus, denn diese Seiten können **nicht** per Copy and Paste übertragen werden. Der Text lässt sich dann nämlich nicht „anfassen“ werden.

Um diese Seiten in ein anderes Format bringen zu können, benötigt man eine OCR (Optical Character Recognition) Software für die Dokumentenverarbeitung.

TTS arbeitet seit längerem mit dem ABBYY FineReader Programm (www.abbyy.com), das sich in der Praxis bislang gut bewährt hat.

Mit Hilfe dieses Programms ist es möglich, Texte aus Graphiken zu erfassen sowie Graphiken und Tabellen ins Wordformat zu übernehmen. Man erhält ein „Rohlayout“, das selbstverständlich noch einer weiteren Bearbeitung bedarf, um es kundengerecht liefern zu können.

Diese Methode ist insbesondere bei aufwendig gestalteten Seiten ein großer Fortschritt für die Erfassung von Texten aus PDF-Formaten. Jedoch ist der Aufwand auch hier nicht zu unterschätzen, da insbesondere die in Graphiken eingebundenen Texte alle einzeln angefasst und bearbeitet werden müssen. Diese erscheinen als Textrahmen bzw. Positionsrahmen im Wordformat und müssen auch dort jeweils einzeln ausgewählt, übersetzt und dann wieder layoutet werden.

Handhabung und Aufwand

Viele Kunden unterschätzen oft den Arbeitsaufwand, der mit PDF-Dokumenten verbunden ist. „Copy and Paste“ reicht leider oftmals nicht aus, um eine nicht nur inhaltlich, sondern auch optisch und sachlich qualitativ hochwertige Arbeit abliefern zu können, die unsere Kunden zu Recht von uns verlangen.

Da oft Dokumente auch in asiatischen Ländern hergestellt werden, dort jedoch andere Betriebssysteme und Programme erforderlich sind, können viele unserer Kunden nur PDF-Formate liefern. TTS ist bemüht, diesen Anforderungen gerecht zu werden. Bitte haben Sie aber Verständnis dafür, dass wir nicht alle Layout-Nachbearbeitungen kostenlos durchführen können. Wir hoffen, Ihnen mit dieser kurzen Beschreibung einen kleinen Einblick in die Weiterverarbeitung von PDF-Formaten gegeben zu haben – sollten Sie weitere Informationen wünschen, rufen Sie uns gerne an.



Allgemeine Informationen zu PDF

Portable Document Format (PDF)

Das **Portable Document Format (PDF)** ist ein plattformübergreifendes Dateiformat für druckbare Dokumente, das von der Firma Adobe Systems entwickelt und 1993 mit Acrobat 1 veröffentlicht wurde. PDF ist ein proprietäres, aber offengelegtes Dateiformat, das im PDF Reference Manual von Adobe dokumentiert ist. Eine Teilmenge des Formats ist inzwischen als PDF/X und PDF/A von der ISO genormt worden.

Damit lassen sich aus Texten, Bildern und Grafiken gemischt oder nur einzeln bestehende Dokumente/Dateien erzeugen (z. B. mit Adobe Acrobat Professional 7.0) und darstellen. Die Benutzenden sollen dieses Format nicht ändern und bearbeiten. Dadurch soll u. a. das Copyright besser gewahrt werden. Etwas daraus zu kopieren, kann beim Erstellen erlaubt oder verboten werden.

In der Startphase war der Adobe Reader kostenpflichtig. Erst die kostenfreie Weitergabe der Software ermöglichte die Verbreitung im heutigen Ausmaß. Die aktuelle Version des PDF-Dateiformats ist 1.6 und kann mit Acrobat ab Version 7 verarbeitet werden. PDF-Dateien geben das mit dem Erstellungsprogramm erzeugte Layout in einer vom Drucker und von Voreinstellungen unabhängigen weitgehend originalgetreuen Darstellung wieder. Ein häufig verwendetes Programm zur Erzeugung von PDF ist Adobe Acrobat Distiller, der aus PostScript-Dateien PDF erstellt. Der Distiller ist verfügbar für Windows und Mac. Diverse Office- und DTP-Anwendungen von Drittherstellern bieten einen direkten PDF-Export an und sind auf vielen Plattformen verfügbar. Mit weiteren Werkzeugen lassen sich – oft über den Druckbefehl, bei größeren Datenmengen aber meist mit einer Programmiersprache aus einer Datenbank heraus – PDF-Dateien einfach und schnell erstellen. Die Erzeugung von PDF-Dateien ist damit auf jeder Plattform möglich.

Durch Offenlegung und Normung von PDF können Drittentwickler unabhängig von Adobe PDF-Werkzeuge bereitstellen. PDF basiert zu großen Teilen auf dem PostScript-Format, das ebenfalls offen gelegt ist.

Verwendung und Vorteile

Eine PDF-Datei gibt die Dokumente des Ursprungsprogramms einschließlich aller Schriften, Farben, Raster- und Vektorgrafiken präzise wieder. Diese Dokumente können eine oder tausende Seiten Umfang haben, wobei jede auf eine maximale Seitengröße von 508 cm x 508 cm beschränkt ist.

Das PDF basiert auf dem gleichen Grafikmodell wie PostScript. Im Gegensatz zu PostScript ist es aber keine Programmiersprache, sondern eine Dokumentenbeschreibungssprache. PDF erlaubt eine genauere Strukturierung von Dokumenten, als das mit Postscript möglich ist. Schriften (außer Pixelschriften) jeglicher Art und Vektorgrafiken können dabei beliebig ohne Qualitätsverlust vergrößert werden. Große Netzwerkpläne und Datenmodelle lassen sich unter diesen Voraussetzungen auf einer PDF-Seite unterbringen.



Aus PDF-Dokumenten lassen sich Textpassagen, Tabellen und Grafiken (und auch Ausschnitte davon) leicht in anderen Anwendungsprogrammen durch Kopieren und Einfügen der jeweiligen Elemente weiterverarbeiten. Text kann nicht nur zur Weiterverarbeitung in anderen Anwendungen, sondern auch zum Durchsuchen oder zur Verwendung mit anderen Ausgabemedien wie beispielsweise Screenreadern extrahiert werden. Durch die Textsuche im einzelnen Dokument oder die Volltextrecherche innerhalb einer PDF-Dokumentensammlung lassen sich sehr einfach Detailinhalte auffinden. Dies funktioniert selbst dann, wenn der Text grafisch verzerrt, z. B. in Kreis- oder Kurvenform dargestellt ist.

Eine Besonderheit des PDFs ist der optionale Dokumentenschutz mit 40- oder 128-Bit-Verschlüsselung. Durch Vergabe eines Benutzerpasswortes ist es so möglich, das Dokument nur einem beschränkten Personenkreis zugänglich zu machen. Weiterhin kann der Autor mit einem separatem Besitzerpasswort gezielt die Rechtevergabe des betreffenden Dokuments festlegen. So kann verhindert werden, dass Benutzer das Dokument abändern, ausdrucken oder Teilinhalte kopieren können. Die Kenntnis des Benutzerpasswortes vorausgesetzt lassen sich diese Rechteeinschränkungen jedoch mithilfe diverser Werkzeuge entfernen.

Durch entsprechende PDF-Werkzeuge lassen sich auch Rechte vergeben, die es ermöglichen, PDF-Dokumente mit Notizen, Kommentaren und Dateianhängen zu versehen oder Formulareinträge abzuspeichern. Ursprünglich konnten diese Merkmale nur mit Adobe Acrobat genutzt werden, seit Version 7 ist es jedoch auch mit dem kostenlosen Adobe Reader möglich, Notizen und Kommentare hinzuzufügen, sofern das entsprechende Dokument vom Verfasser mit den notwendigen Berechtigungen versehen wurde.

Ursprünglich war das PDF ein reines Ausgabe- bzw. Druckformat. Später gab es auch Programme, mit denen man fertige PDF-Dateien verändern kann, zum Beispiel mit dem Adobe Designer. Das Format ist nicht mit den internen Dateiformaten von Textverarbeitungsprogrammen vergleichbar und eignet sich, abgesehen von der Notiz- und Kommentarfunktion, nur begrenzt zur Weiterverarbeitung von Dokumenten. Es ist allerdings innerhalb gewisser Grenzen möglich, beispielsweise Tippfehler zu entfernen. Vorteile im Desktop Publishing sind für Grafiker und Designer die Einbindung aller Elemente für die Druckerstellung.

PDF-Dokumente können abhängig vom Einzelfall sowohl größer als auch kleiner als die Dateien der Ursprungsanwendung sein. Die Größe eines Dokuments hängt von der Art der enthaltenen Daten, von der Effizienz des Erstellungsprogramms und davon ab, ob Schriftarten eingebettet wurden. Schriften können entweder vollständig, als Untermenge der tatsächlich im Dokument verwendeten Zeichen oder aber überhaupt nicht eingebettet werden. Soll ein Dokument unabhängig davon, ob auf der Zielplattform die verwendeten Schriften installiert sind, zuverlässig darstellbar sein, müssen mindestens die tatsächlich verwendeten Zeichen eingebettet werden.



Das PDF wurde im Laufe seiner Entwicklung mehrfach auf spezielle Anforderungen für die Verwendung im Internet angepasst. So musste ein Dokument ursprünglich vollständig verfügbar sein, um dargestellt werden zu können. Inzwischen ist es möglich, PDF-Dokumente zu linearisieren, so dass bereits Teile eines Dokuments während des Ladevorgangs dargestellt werden können. Seit der Version 1.5 der PDF-Spezifikation kann der größte Teil der Dokumentstruktur komprimiert werden.

Heute ist es möglich, interaktive PDF-Formulare zu gestalten. In ihnen kann man in bestimmten Eingabefeldern Werte eingeben.

Es ist auch möglich, Dokumente als Bild einzuscannen und zusätzlich die entsprechenden Informationen als durchsuchbaren Text im gleichen PDF-Dokument zu erzeugen.

Eigenschaften von PDF-Dateien

Eine der ureigensten Stärken des PDFs ist die plattformunabhängige immer gleiche Darstellung des Inhaltes. Dies bedeutet, dass auf jeder Hard- und Softwareplattform der Inhalt eines PDFs ohne graphischen Unterschied dargestellt wird. PDF ist daher auch ein Format zum Publizieren von Inhalten. Während Vorarbeiten oft in Text-Editoren gemacht werden, wird das fertige Werk dann als PDF veröffentlicht. Dies wird in der Wissenschaft, Wirtschaft, aber auch von staatlicher Seite gemacht (z. B. Bundesgesetzblätter).

Da PDFs oft für das Publizieren von urheberrechtlich geschützten Werken verwendet werden, ist es üblich, das Werk entweder als im PDF zusammengefasste nicht als Text zu bearbeitende Bilddateien oder als gesperrtes (Druck- und Editierfunktion im AcrobatReader ausgeschaltet) bzw. verschlüsseltes PDF bereitzustellen. Insbesondere die bei wissenschaftlichen Datenbanken sehr beliebte erste Methode führt dazu, dass die PDFs im Vergleich zu reinem Fließtext merkbar größer sind.

Sicherheitslücken durch menschliches Versagen

Im Internet publizierte PDF-Dateien können unbeabsichtigt vertrauliche Informationen offenlegen. Die übersehenen Peinlichkeiten unterlaufen meist beim Konvertieren von Textdateien nach PDF.

Vermeidung

Es ist relativ einfach, sich gegen die Weitergabe vertraulicher Informationen abzusichern. Das Abspeichern der ursprünglichen Textdatei in eine neue Datei löscht in der Regel die im Original enthaltene Bearbeitungschronik. Nach Umwandlung dieser neuen Textdatei nach PDF ist zusätzlich eine sorgfältige Prüfung der in der PDF abgespeicherten Dateieigenschaften erforderlich.



PDF in Betriebssystemen

Das Einsatzgebiet von PDF ist vielfältig. So ist es nicht verwunderlich, dass mit dem Betriebssystem Mac OS X von Apple erstmals PDF als Standardformat für die Bildschirmausgabe verwendet wurde. Die Vorteile liegen auf der Hand: Mit Hilfe von PDF ist erstmals echtes WYSIWYG möglich. Weitere Features sind Textglättung in allen Anwendungen sowie die PDF-Erzeugung aus jeder Anwendung heraus, die einen Druckdialog besitzt. PDF wird auch zur Erzeugung der Druckdaten verwendet, damit ist es möglich, PostScript auf Nicht-Postscript-Druckern auszugeben.

Unter Windows wie auch unter GNU/Linux gibt es den kostenlosen Acrobat Reader, um PDF-Dateien anzuschauen. Für unixbasierte Systeme gibt es zusätzlich das äußerst schnelle Programm XPDF, welches auf die grundlegendsten Funktionen (Anzeige auf Bildschirm, Durchsuchen des Dokuments, Ausdrucken) reduziert ist, und deshalb viel schneller geladen und gestartet werden kann und einige weitere Programme (KPDF und Evince), die an ihre Desktop-Umgebung angepasst sind.

Boxen im PDF

Die verschiedenen Boxen in einem PDF-Dokument sind vor allem in der Druckindustrie sehr wichtig, da fehlerhafte PDF-Dateien nicht korrekt gedruckt werden können. In der Druckindustrie wird meist der Adobe Acrobat Distiller für die PDF-Erzeugung eingesetzt, da dieser ein standardkonformes PDF-Dokument erzeugt.

MediaBox

Sie definiert die Größe des Ausgabemediums und den *Medienrahmen* des PDF-Dokumentes. Das Dokument ist noch nicht beschnitten und enthält oft die im PDF-Generator eingestellte PostScript-Seitengröße. Die Media Box muss immer die größte aller Boxen sein, da sie alle anderen (nachfolgend erläuterten) Boxen mit einschließen muss.

BleedBox

Eine Bleedbox beinhaltet Informationen über die *Anschnittrahmen*, die die Größe des Endformates mit einem Beschnitt definiert. In der Druckindustrie wird ein Beschnitt von minimal 3 mm pro Seite benötigt. Ein Anwendungsbeispiel sind Bilder, die angeschnitten werden sollen und somit aus dem Rand des Drucks laufen. Für eine A4 Seite, die an eine Druckerei geliefert werden soll, ergibt sich somit eine Breite von 210 mm + 6 mm und eine Länge von 297 mm + 6 mm (A4 extra mit 216 mm × 303 mm). Dies muss zwingend eingehalten werden, um keine Blitzer zu erzeugen, da die Druckmaschine etwas variieren kann.



TrimBox

Die TrimBox ist das Endformat einer PDF-Datei ohne Beschnitt und wird auch *Endformatrahmen* genannt. In der Software Adobe Acrobat wird diese als *Crop-Box* oder *Maskenrahmen* bezeichnet, und beinhaltet ebenfalls das beschnittene Format; bei der *Crop-Box* sind jedoch noch die Informationen, die sich noch außerhalb der Box befinden, vorhanden und werden lediglich ausgeblendet.

Art Box (auch Bounding Box)

Dies stellt den *Objektrahmen* dar, das heißt, diese Box ist die kleinstmögliche Box, die alle Objekte einschließt, die sich auf der PDF-Seite befinden. Sie gibt auch die Größe der Grafik beim Import einer Datei (beispielsweise EPS) an.

PDF in der Druckvorstufe

Ziel der Boxinformationen ist es, ein PDF in der Druckvorstufe besser weiterverarbeiten zu können. Heute müssen viele PDF noch manuell nachbearbeitet werden, um sie auch in einer Druckmaschine drucken zu können, da nur wenige Programme die Informationen gut einbinden (beispielsweise InDesign oder QuarkXPress). Als Hilfsmittel in der Druckvorstufe sind unter anderem die Softwareprodukte Asura(Preflight), Speedflow, PitStop Professional, Prinergy, Prinect Printready und Puzzleflow verbreitet, welche PDF-Dateien für den Druck aufbereiten können, Preflight genannt.

Programme zur Erstellung von PDF für die Druckvorstufe

Es führen zwei Wege zur Erstellung: Entweder durch die Verwendung eines Druckertreibers auf PostScript-Basis, oder direkte PDF-Erzeugung, bei der die PostScript-Daten unsichtbar sind beziehungsweise nur temporär benötigt werden.



Programme zum Betrachten von PDF-Dateien

Freeware-Programme

Adobe Reader / Acrobat Reader (Windows, Mac OS, UNIX)

Foxit PDF Reader (Windows)

GSview (Windows, OS/2)

eXPert PDF Reader (Windows)

CAD-KAS PDF Reader (Windows)

Brava! Reader

Apple Vorschau (engl. Preview) (Mac OS X) - integraler Bestandteil des Betriebssystems

Open-Source-Programme

Xpdf (X11)

evince und Gpdf (GNOME, basierend auf Xpdf)

Kpdf (KDE, basierend auf Xpdf)

Ghostview, gv (X11) in Zusammenarbeit mit Ghostscript

GGv (GNOME, basierend auf Ghostscript)

KGhostview (KDE, basierend auf Ghostscript)

PalmPDF (Palm OS 5.x, basierend auf Xpdf)

Systeminterne Dienstprogramme

Vorschau (Mac OS X) – integraler Bestandteil des Betriebssystems

PDFViewer (PC/GEOS und DOS)

Online-Betrachtung (keine Installation nötig)

Online-Viewer für *.pdf, *.ps und *.doc

Programme zum Erstellen von PDF-Dateien

Open Source:

StarOffice und OpenOffice.org können jedes Dokument als PDF speichern.
pdfTeX, pdfLaTeX sind Ergänzungen zu LaTeX, die fast alle PDF-Fähigkeiten unterstützen.

KOffice ist ein freies Office-Paket mit PDF-Unterstützung.

Scribus

xfig ist ein Vektor-Graphik-Programm, das als „.pdf“ abspeichern kann.

Freeware:

602pc Suite Freeware zum Erstellen von PDF

Ghostscript, ein Interpreter für Postscript-Dateien

FreePDF – eine grafische Benutzeroberfläche für Ghostscript unter Windows-Betriebssystemen (deutsch)



Kostenpflichtige Software:

Adobe Acrobat Distiller (aktuelle Versions-Nr. 7.0), Adobe Illustrator, Adobe Photoshop, Adobe InDesign, Adobe Elements, Freehand
Mac OS X (im aktuellen Betriebssystem von Apple kann aus jedem druckfähigen Programm ein PDF erzeugt werden)
pdf-Office speziell für PDF-Formulare
CorelDRAW, Corel WordPerfect
CATIA CAD-Programm mit PDF-Unterstützung
Jaws PDF-Creator
Abbyy FineReader
QuarkXPress
PDFlib Bibliothek zur Erstellung eines PDFs für C, C++, Java und diverse
Scriptsprachen

Achtung: Dient das PDF als Vorlage für einen Druck in einer Druckerei, sollte diese vor der Erzeugung des PDFs kontaktiert werden, da einige der oben angeführten Programme Probleme in der Druckvorstufe bereiten.