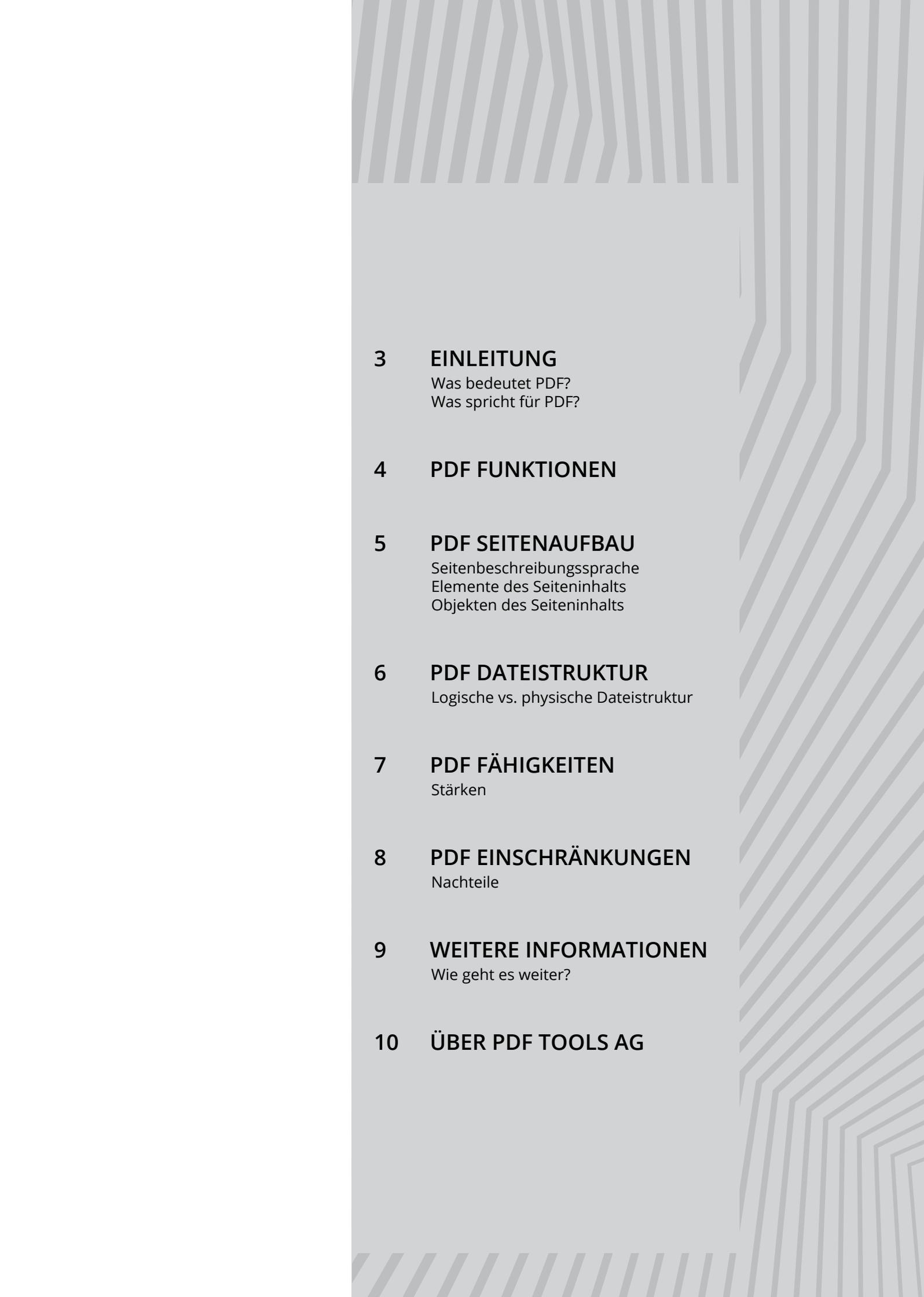




WHITEPAPER

PDF GRUNDLAGEN

Einführung in das Format PDF

- 
- 3 EINLEITUNG**
Was bedeutet PDF?
Was spricht für PDF?
 - 4 PDF FUNKTIONEN**
 - 5 PDF SEITENAUFBAU**
Seitenbeschreibungssprache
Elemente des Seiteninhalts
Objekten des Seiteninhalts
 - 6 PDF DATEISTRUKTUR**
Logische vs. physische Dateistruktur
 - 7 PDF FÄHIGKEITEN**
Stärken
 - 8 PDF EINSCHRÄNKUNGEN**
Nachteile
 - 9 WEITERE INFORMATIONEN**
Wie geht es weiter?
 - 10 ÜBER PDF TOOLS AG**

EINLEITUNG

PDF Dokumente sind weltweit verbreitet und in fast allen Marktsegmenten präsent. Die meisten unter uns kennen und verwenden den Ausdruck „PDF“, haben aber eine ungenaue Vorstellung davon, was das Dateiformat kann und was nicht. Dieses White Paper erklärt, was PDF wirklich ist und wo die Stärken und Schwächen liegen.

Was bedeutet PDF?

Die Abkürzung PDF steht für „Portable Document Format“ und bezeichnet ein Dateiformat. PDF wurde ursprünglich von Adobe Systems Inc. entwickelt und ist seit 2008 ein ISO Standard.

Warum wurde PDF entwickelt?

PDF wurde in den frühen 1990er Jahren als Plattform unabhängiges Dateiformat - mit folgenden Zielen – kreiert :

- Austausch und Darstellung elektronischer Dokumente ermöglichen
- Texte und Bilder unabhängig von der Auflösung graphisch darstellen
- Dokumente für die (Web-) Ansicht optimieren
- Interaktive Funktionen anbieten

Was spricht für PDF?

PDF ist als elektronisches Dokumentenformat aus mehreren Gründen attraktiv:

- **Portabilität** - PDF ist Plattform unabhängig; so lässt sich z.B. eine PDF Datei, die mit einer Windows Applikation erstellt wurde, anschliessend auf einem Linux Server weiterverarbeiten und auf Macintosh Computern betrachten.
- **Elektronisches Dokument mit vielen Extrafunktionen** - PDF baut auf der äusserst erfolgreichen PostScript Seitenbeschreibungssprache auf, und bietet viele Extrafunktionen: Direkter Zugriff auf Seiten, Komprimierung, Verschlüsselung, interaktive Navigation, Formulare uvm.
- **Industriestandard** – PDF hat sich zum international anerkannten ISO-Standard (ISO 32000) für den elektronischen Austausch von Dokumenten entwickelt. Zudem ist PDF heute das am meisten eingesetzte Format für die Produktion von Druckvorlagen in der digitalen Druckvorstufe. Private Organisationen, Behörden und das Bildungswesen überarbeiten ihre Businessprozesse, indem sie Papier basierte Arbeitsprozesse mit elektronischem Informationsaustausch ersetzen.
- **Kostenloser Betrachter** - Einer der Hauptgründe, warum sich PDF derart weit verbreitet hat, ist der PDF Reader von Adobe. Dieser ist seit langem kostenlos erhältlich. Bezahlen muss man nur für diejenigen Applikationen, mit denen man PDF Dokumente herstellt oder bearbeitet.

PDF FUNKTIONEN

PDF bietet mehrere Funktionen an, die das Format vielseitig und manchmal einzigartig machen:

- **Bild und Darstellung sind getrennt** - unterscheidet zwischen Graphik (Formen und Farben) und Darstellung (Rasterung im Ausgabegerät). Die Darstellung der Seiten kann im PDF Dokument in einer geräteunabhängigen Art und Weise spezifiziert werden. Die Wiedergabe der Seiten (z.B. zum Drucken oder Betrachten) kann auf der Basis der spezifischen Eigenschaften der Ausgabegeräte optimiert werden.
- **Komprimierung** - PDF Objekte, insbesondere Bilder, werden dank unterschiedlichen Komprimierungsalgorithmen sehr effizient und ohne sichtbaren Qualitätsverlust komprimiert. Deshalb kann ein PDF Dokument einen Bruchteil der Dateigrösse des Originals ausmachen.
- **Schriftenverwaltung** - Alle im PDF verwendeten Schriften können direkt im Dokument eingebettet werden. Das garantiert, dass der Text bei der Darstellung immer exakt gleich aussieht. Um Platz zu sparen, lassen sich die Schriften einschränken, sodass sie nur noch über diejenigen Zeichen verfügen, die auch wirklich verwendet werden.
- **Erzeugung in einem Arbeitsschritt** - Wenn ein PDF erzeugt wird, ist die physische Anordnung der Objekte im Dokument irrelevant. Die Objekte müssen nicht zuerst in einem Vorbereitungsschritt geordnet werden. Das erlaubt es, ein PDF in einem einzigen Arbeitsschritt zu erstellen.
- **Direktzugriff** - man kann auf die Objekte innerhalb einer PDF Datei direkt zugreifen. Wenn man beispielsweise die Seite 733 eines 800-seitigen Dokuments aufruft, identifiziert PDF zuerst diejenigen Objekte, die für die Darstellung dieser Seite benötigt werden. Man muss nicht warten, bis das gesamte Dokument geladen wird.
- **Sicherheit, Verschlüsselung, digitale Signaturen** - PDF unterstützt verschiedene Arten von Verschlüsselungen, Zugangskontrollen und digitalen Signaturen. Deshalb ist das Format besonders geeignet für die Bearbeitung vertraulicher Dokumente, die übers Internet verschickt oder in einer Webbrowserapplikation verwendet werden. PDF Dokumente lassen sich so verschlüsseln, dass niemand den Inhalt ohne Passwort rekonstruieren kann.
- **Erweiternde Aktualisierung** - wenn ein PDF Dokument aktualisiert wird, werden die abgeänderten Objekte einfach am Schluss des Dokuments hinzugefügt. Das PDF muss nicht von Grund auf neu erzeugt werden. Deshalb können kleinere Ergänzungen (z.B. Wasserzeichen hinzufügen oder Textkorrekturen) schnell und einfach angebracht werden. Grössere Änderungen können jedoch auf diese Weise zu einer grösseren Datei führen, jedoch jederzeit in einem Optimierungsprozess konsolidiert werden.
- **Praktische Funktionen für Dokumentenaustausch** - Dokumente verfügen über zahlreiche Funktionen, welche das endgültige Aussehen nicht beeinflussen, aber dennoch sehr nützlich sind für den Austausch von. Ein Beispiel dafür sind die eingebetteten Metadaten (Titel, Autor, Datum der Erstellung oder Modifikation, etc.).

PDF SEITENAUFBAU

Seitenbeschreibungssprache

Die zentrale Funktion eines PDF ist die Seitenbeschreibungssprache. Sie legt das Aussehen einer Seite solchermaßen fest, dass sie für den Druck oder zum Betrachten konsistent reproduziert werden kann. Die Sprache ist eine Weiterentwicklung des PostScript-Modells. Um effiziente Prozesse zu ermöglichen, ist sie allerdings einfacher umzusetzen: die Sprache enthält beispielsweise keine Programmieranweisungen wie Schleifen oder „wenn..dann“ Befehle.

Elemente des Seiteninhalts

Grundsätzlich erkennt PDF drei Arten von Seiteninhaltelementen:

- Text
- Grafische Pfade (Linien und Kurven)
- Bilder

Objekten des Seiteninhalts

Die Seitenbeschreibungssprache von PDF verwendet Objekte, um den Inhalt zu beschreiben. Jede Textzeile und sämtliche Graphiken und Bilder werden durch eine Folge von Objekten definiert (die selbst wiederum durch eine Folge von Operatoren beschrieben sind).

- **Text** – Textobjekte werden durch folgende Attribute bestimmt: Schriftart, Schriftgröße, Zeichenfolge, Farbe, die Position auf der Seite und die Art der Darstellung wie das Ausfüllen, Umranden usw. Informationen wie Zeilenumbruch, Kopfzeile, Paragraph, Einzug usw. sind darin nicht enthalten. Das heisst, dass wichtige Hinweise zur Formatierung von Paragraphen, wie sie z.B. im Textverarbeitungsprogramm Microsoft Word verwendet werden, nicht gespeichert werden. Der eigentliche Text wird in Fragmente aufgeteilt, die so klein wie einzelne Zeichen oder so gross wie eine ganze Zeile sein können. Diese Fragmente werden zufällig gespeichert und sind wie Puzzleteile, die erst durch die richtige Positionierung ein Gesamtbild ergeben.
- **Graphikpfade** – Ein graphischer Pfad ist eine frei wählbare Form, die aus Elementen wie geraden Linien, Rechtecken und Kurven (Bézier, kubisch) besteht. Die Elemente können frei sein, beliebig zusammenhängen und sich schneiden. So definierte Pfade können als Linien gezeichnet, gefüllt oder zum Ausschneiden (Clip) anderer Inhalte verwendet werden. Die Darstellung eines Pfades kann durch zahlreiche Attribute wie Farbe, Linienbreite uvm. gestaltet werden.
- **Rasterbilder** – Bilder werden nicht im Originalformat eingebettet. PDF ist unabhängig von der ursprünglichen Bilddatei. Im TIFF Format gespeicherte Scans oder die Segmente von JPEG-Bildern, die in ein PDF umgewandelt werden, werden im eigenen PDF Bildformat gespeichert. Danach ist es in der Regel nicht mehr möglich, das Originalformat zu bestimmen. Allerdings kann man PDF Bilder jederzeit in gerasterte Bildformate exportieren, wenn das gewählte Format keine Einschränkungen bezüglich der Bildeigenschaften aufweist.

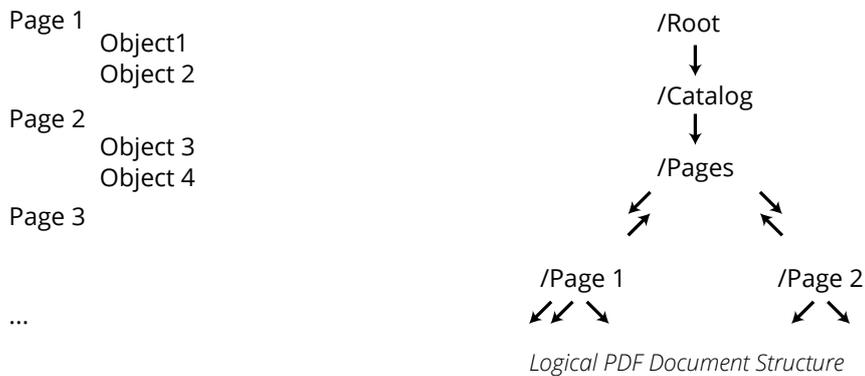
PDF DATEISTRUKTUR

Logische vs. physische Dateistruktur

Wie sieht ein PDF Dokument im Detail aus? Um diese Frage zu beantworten, muss man zwischen logischer Dokumentenstruktur und physischer Dateistruktur unterscheiden.

Logische Dokumentenstruktur

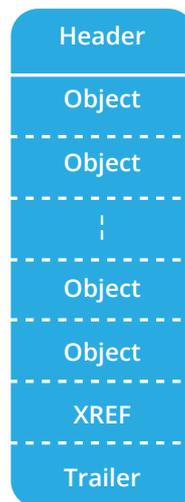
Die logische Struktur eines PDF-Dokuments gibt vor, wie das Dokument beim Betrachten oder für den Druck reproduziert werden soll:



Physische Dateistruktur

Die physische Dateistruktur unterscheidet sich markant von der logischen Dokumentenstruktur. Ein PDF besteht aus einem Dateikopf, aus einer Folge von Objekten, einer Querverweistabelle und einem Nachspann.

Die Objekte auf einer Seite werden nicht in einer bestimmten Reihenfolge gespeichert (z.B. von oben nach unten). Ihr Platz in der physischen Datei ist rein zufällig.



Physical PDF File Structure

Lesen einer PDF Datei

Beim Lesen einer PDF Datei (z.B. für die Darstellung am Bildschirm oder den Druck) wird die Kopfzeile zuerst gelesen, um die Datei als PDF zu identifizieren. Danach wird der Nachspann gelesen. Dieser verweist zur Querverweistabelle, mit deren Hilfe die Objekte gefunden werden, welche die Seiten, Schriften, Bilder usw. enthalten.

PDF FÄHIGKEITEN

Stärken

PDF bietet viele Vorteile. Diese Stärken haben wesentlich zum Erfolg und zur Verbreitung des Formats beigetragen.

- PDF ist ein weltweit anerkannter Industriestandard. PDF ist seit 2008 von der ISO (International Standards Organisation) als Standard mit der Nummer ISO-32000 freigegeben.
- PDF wurde auch im Hinblick auf die wachsende Bedeutung des World Wide Web konzipiert. Während der rasanten Verbreitung des Webs in den vergangenen zehn Jahren hat sich PDF als Standardformat etabliert. Hauptsächlich für die elektronische Bereitstellung von Dokumenten aller Art
- Mehrseitige PDF Dokumente lassen sich für eine schnelle Webansicht optimieren. Man kann auch in riesigen Dokumenten mit mehreren tausend Seiten eine bestimmte Seite direkt abrufen bzw. anschauen, bevor die restlichen Seiten geladen werden.
- PDF verfügt über sehr leistungsfähige Komprimierungsverfahren. Vor allem Bilder lassen sich damit ohne sichtbaren Qualitätsverlust und mit bedeutend geringerem Datenvolumen speichern.
- Der PDF Betrachter von Adobe ist immer noch kostenlos erhältlich. Unterdessen gibt es aber auch zahlreiche Betrachter von anderen Herstellern sie beispielsweise der Betrachter von PDF Tools AG.
- PDF unterstützt mehrseitige Dokumente beliebiger Seitenformate.
- Adobe hat eine herausragende PDF Spezifikation kreiert. Sie ist sehr ausführlich und detailliert und beinhaltet explizite Details über sämtliche Aspekte des Formats. Das erlaubt den Softwareunternehmen, eigene Programme zu schreiben, die PDF Dokumente erzeugen und lesen können.
- PDF eignet sich für viele Einsatzbereiche. Der Markt bietet eine vielfältige Palette an Funktionen an, mit welchen man PDF Dokumente bearbeiten kann: in einem Arbeitsschritt erstellen, zusammenfügen, trennen, mit Stempel versehen, Inhalt extrahieren, verschlüsseln, konvertieren, betrachten, drucken, mit Formularverarbeitung verwenden usw.
- PDF ist ohne Frage aus dem Status eines Dokumentformats herausgewachsen. Gängige Technologien wie Multimedia, XML, Java, Formularverarbeitung, Komprimierung, individuelle Verschlüsselung usw. sind Teil der Spezifikation geworden. Mit PDF hat man ein leistungsstarkes und vielseitiges Format für intelligente Anwendungen zur Verfügung.

PDF EINSCHRÄNKUNGEN

Nachteile

Das PDF Format beinhaltet trotz der vielen Vorteile jedoch auch gewisse Einschränkungen.

- **Kompatibilitätsprobleme:** Es gibt PDF-Dateien, die mehr als zehn Jahre alt sind. Die Erfahrung hat gezeigt, dass gewisse ältere Dateien mit neueren Betrachtern nicht mehr geöffnet werden können. Umgekehrt verhindert die rasante Entwicklung neuer Funktionen die Anzeige auf älterer Software und zwingt den Anwender zu laufenden Upgrades.
- Das PDF-Format ist sehr komplex und wird nicht von allen Software-Herstellern korrekt implementiert. Durch qualitativ ungenügende Werkzeuge erstellte Dateien haben Konformitätsmängel, d.h. Probleme bezüglich der Übereinstimmung mit der Spezifikation. Das Problem wird noch dadurch verschärft, dass die Adobe Reader zu tolerant sind und der Anwender deshalb die Mängel nicht oder zu spät entdeckt
- PDF wuchert und beginnt, zu viele Fremdformate bzw. -technologien für eingebettete Objekte anzubieten (JavaScript, XMP, XFA, PJTF uvm.). Jedes einzelne dieser Formate ist in sich sehr komplex und entsprechend schwer zu beherrschen. Für sie gilt das oben gesagte in besonderem Masse.
- PDF folgt nicht zwingend dem Prinzip WYSIWYG (what you see is what you get). Das betrifft vor allem die Darstellung von Farben und Schriften. PDF Dokumente können sich bei der Präsentation von der ursprünglichen Absicht unterscheiden. Es gibt jedoch Standards wie der im Jahre 2005 veröffentlichte ISO-Standard PDF/A, die sich speziell diesen Problemen annehmen und sie durch gezielte Gebote und Verbote verhindern.
- Die Seitenbeschreibungssprache wurde primär für die Graphikindustrie entworfen und nicht für Anwendungen wie Textverarbeitungen. PDF kennt zwar ergänzende Attribute wie die Dokumentenstruktur, die Leserichtung und das Erläutern von Bildern durch alternative Texte. Diese Eigenschaften sind jedoch optional, schwierig zu erstellen und zu handhaben. Entsprechend erzeugte Dateien (sog. „tagged“ PDF) kommen deshalb immer noch selten vor.

WEITERE INFORMATIONEN

Wie geht es weiter?

Das Ziel dieses White Paper ist eine grundlegende Einführung in die Welt von PDF. Hoffentlich haben Sie nun eine bessere Vorstellung davon, was PDF ist, wie es strukturiert ist, und was die wichtigsten Stärken und Schwächen sind.

Weitere White Paper

Wenn Sie mehr über spezifische PDF Technologien wissen wollen, unterstützen Sie unzählige Portale und White Paper auf dem Internet. PDF Tools AG (<http://www.pdf-tools.com>) publiziert eine Sammlung von White Paper, die sich mit einer Vielzahl von PDF Technologien beschäftigen

PDF Expert Blog

Für Informationen von den PDF-Experten aus unserem Entwicklungsteam, besuchen Sie unseren Blog unter blog.pdf-tools.com.

Wir sind für Sie da

Wenn Sie weitere Informationen über Normen, Vergleiche und Produktinformationen einschließlich eines auf Ihre Bedürfnisse zugeschnittenen Angebots wünschen, zögern Sie nicht, uns zu kontaktieren.

Komponenten & Lösungen für PDF and PDF/A Dokumentenprozesse

Holen Sie sich Ihre eigene voll funktionsfähige Testversion unserer PDF-Software für 30 Tage. Besuchen Sie unsere Website für weitere Informationen.

ÜBER PDF TOOLS AG

Die PDF Tools AG zählt weit über 4'000 Unternehmen und Organisationen in 60 Ländern zu ihren Kunden und ist damit ein weltweit führender Hersteller von Softwarelösungen und Programmierkomponenten für PDF und PDF/A Produkte.

Vor mehr als 15 Jahren hat Dr. Hans Bär-fuss, der Gründer und CEO der PDF Tools AG, PDF Technologie in Kundenprojekten eingesetzt. Das PDF und PDF/A Format haben sich seither zu einem mächtigen, weit verbreiteten Format und ISO Standard mit nahezu unbeschränkten Anwendungsmöglichkeiten entwickelt. Auch die PDF Tools AG wurde in dieser Zeit zu einem der wichtigsten Unternehmen im Markt für PDF Technologie und hat den PDF/A ISO Standard für die elektronische Langzeitarchivierung massgeblich mitgeprägt.

Als Schweizer Vertreter im ISO-Komitee für PDF/A und PDF, lässt das Unternehmen sein Wissen direkt in die Produktentwicklung einfließen. Es entstehen damit qualitativ hochwertige Produkte mit einer effizienten Leistungsfähigkeit – ganz nach dem 3-Heights™ Motto des Entwicklungsteams, welches sich aus qualifizierten Ingenieuren zusammensetzt.

Das Portfolio von PDF Tools AG erstreckt sich von Komponenten, Services bis hin zu Lösungen. Die Produkte unterstützen den gesamten Dokumentenfluss vom Rohmaterial, über Scanningprozesse bis hin zur Signierung und der Archivierung in einem rechtlich konformen Langzeit-

archiv. Ein Vorteil der Komponenten und Lösungen ist die breite Palette von Schnittstellen, welche eine reibungslose und einfache Integration in bestehende Umgebungen gewährleistet.

Laufend werden die Produkte aufgrund der wachsenden Anforderungen des Marktes weiterentwickelt. Da der Support von den Entwicklern selbst übernommen wird, erkennen diese Trends und Bedürfnisse der Kunden rasch und setzen das Wissen bei der Planung für Erweiterungen und Komponenten ein.

Sämtliche Entwicklungen erfolgen innerhalb der PDF Tools AG in der Schweiz. Das Unternehmen lagert bewusst keine Programmierung aus, um den gesamten Entwicklungsprozess zentral an einem Standort zu haben. Dies um den eigenen Ansprüchen zu der 3-Heights™-Technologie und der Erfüllung des Firmen-Credos gerecht zu werden.

Der Erfolg der Produkte im Markt bestätigt PDF Tools AG diesen Ansatz. Zu den Kunden des Unternehmens zählen namhafte, weltweit tätige Unternehmen aus sämtlichen Branchen. Das ist das schönste Kompliment und die beste Motivation für das Team, einen wichtigen Beitrag in der PDF und PDF/A Welt zu leisten.





The 3-Heights™ Produktfamilie von PDF Tools AG steht für:

High Quality – High Volume – High Performance

Copyright © PDF Tools AG. All rights reserved.

Namen und Marken Dritter gelten als rechtlich geschütztes Eigentum. Die Rechte können jederzeit geltend gemacht werden. Die Darstellung von Produkten und Dienstleistungen Dritter dienen ausschliesslich zu Informationszwecken.

PDF Tool AG ist für die Performance und den Support von Produkten von Drittfirmen nicht verantwortlich und übernimmt keine Gewähr bezüglich Qualität, Zuverlässigkeit, Funktionalität und Kompatibilität dieser Produkte und Geräte.

PDF Tools AG
Kasernenstrasse 1, 8184 Bachenbülach, Schweiz
Tel.: +41 43 411 44 51 , www.pdf-tools.com