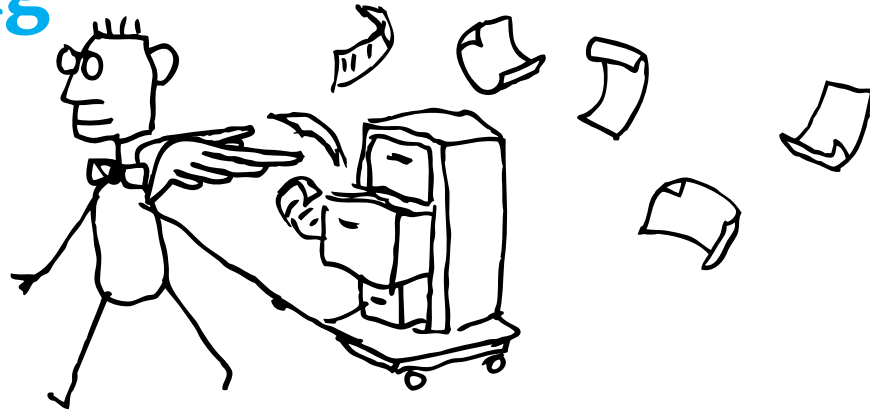


PDF/A – eine technische Einführung



 PDFlib® Whitepaper

Die PDF/A-Archivierungsstandards

PDF/A ist auf die zuverlässige Langzeitarchivierung digitaler Dokumente mit Text, Rasterbildern und Vektorgrafik sowie den zugehörigen Metadaten ausgerichtet. Das in der ISO-Norm 19005 spezifizierte PDF/A-Format soll eine konsistente und stabile Teilmenge von PDF bieten, die sich auch nach langer Aufbewahrungsfrist noch eindeutig reproduzieren oder für den zuverlässigen Datenaustausch in Unternehmen und Behörden verwenden lässt. Dieses Whitepaper beschreibt wichtige technische Aspekte von PDF/A-1, PDF/A-2 und PDF/A-3.

PDF/A-1 PDF/A-1, der erste von mehreren Teilen des Standards, wurde im Jahr 2005 als ISO 19005-1 veröffentlicht. Er basiert auf PDF 1.4 (dem Dateiformat von Acrobat 5) und schreibt gewisse Einschränkungen bei der Verwendung von Farben, Fonts, Anmerkungen und anderen Elementen vor. Es gibt zwei Varianten von PDF/A-1:

- ▶ Konformität zu Level B (PDF/A-1b; »b« wie in »basic«) gewährleistet, dass das Erscheinungsbild eines Dokuments langfristig erhalten werden kann. PDF/A-1b stellt sicher, dass das Dokument unverändert aussieht, wenn es in Zukunft verarbeitet wird.
- ▶ Konformität zu Level A (PDF/A-1a; »a« wie in »accessible« (zugänglich)) basiert auf Level-B-Konformität, hinzu kommen aber wesentliche Eigenschaften der Variante »Tagged PDF«: Gefordert werden Strukturinformationen und zuverlässige Unicode-Textsemantik, um die logische Dokumentstruktur und natürliche Lesereihenfolge zu erhalten. Einfach ausgedrückt gewährleistet PDF/A-1a nicht nur, dass das Dokument bei späterer Verarbeitung unverändert aussieht, sondern auch, dass sein Inhalt (Semantik) zuverlässig interpretierbar und körperlich beeinträchtigten Benutzern zugänglich ist. Als wichtiges Beispiel seien hier Screenreader-Programme erwähnt, die blinden Benutzern »Tagged PDF« vorlesen können.

PDF/A-2 Die PDF-Welt ist seit der Veröffentlichung von PDF/A-1 weit vorangeschritten. Neben vielen anderen Meilensteinen wurde im Jahr 2008 PDF 1.7 (das Dateiformat von Acrobat 8) als ISO 32000-1 standardisiert. Um die zahlreichen neuen PDF-Funktionen in PDF/A-Workflows zugänglich zu machen, wurde im Jahr 2011 PDF/A-2 als neuer Teil des Standard unter der Bezeichnung ISO 19005-2 veröffentlicht.

PDF/A-2 basiert auf PDF 1.7 und enthält viele nützliche, in PDF/A-1 noch nicht verfügbare Neuerungen, wie zum Beispiel wichtige Aspekte des Dateiformats wie die JPEG-2000-Kompression, optionale Inhalte (Ebenen), PDF-Pakete und anderes. PDF/A-2-Dokumente können Dateianhänge enthalten, sofern diese selbst PDF/A-1- oder PDF/A-2-konform sind.

Ähnlich wie bei PDF/A-1 gibt es auch bei PDF/A-2 die Konformitätsstufen B and A. Als neue Variante wurde die Konformität zu Level U eingeführt. Level U liegt zwischen PDF/A-2a und PDF/A-2b: es werden zuverlässige Unicode-Textsemantik, aber keine Strukturinformationen verlangt. PDF/A-2u gewährleistet die eindeutige visuelle Reproduzierbarkeit einer Seite sowie die Möglichkeit, Text zu durchsuchen und zu extrahieren.

PDF/A-2 macht PDF/A-1 nicht obsolet; Benutzer müssen auch nicht auf die neuere Version migrieren – dies wäre wohl absurd bei einem Standard, der auf eine langfristige Erhaltung ausgerichtet ist.

PDF/A-3

Im Jahr 2012 wurde PDF/A-3, ein weiterer Teil der Norm, als ISO 19005-3 veröffentlicht. PDF/A-3 unterstützt auch die Konformitätsstufen A, B und U. Es unterscheidet sich von PDF/A-2 in den folgenden Aspekten:

- ▶ Während PDF/A-2 nur PDF/A-konforme Dateianhänge erlaubt, dürfen bei PDF/A-3 beliebige Dateitypen als Anlagen verwendet werden, um den Anforderungen verschiedener Benutzergruppen gerecht zu werden.
- ▶ Dateianhänge werden dem gesamten Dokument, einer Seite oder einem anderen Teil des Dokuments zugeordnet. Die Beziehung zwischen der angehängten Datei und dem entsprechenden Teil des Dokuments muss explizit angegeben werden, z.B. Quelle, alternative oder ergänzende Daten.

Typische PDF/A-3-Szenarien umfassen die Einbettung von Quelldateien für Textverarbeitungs- oder Tabellenkalkulationsprogramme in einem finalen PDF/A-Dokument oder die Aufnahme von maschinenlesbaren XML-Daten in ein für den menschlichen Gebrauch bestimmtes PDF, z.B. eine Rechnung.

PDF/A-Viewer müssen nichts Besonderes mit den angehängten Nicht-PDF/A-Dateien machen, außer sie zu extrahieren. Der PDF/A-Standard gewährleistet nicht, dass die Anhänge in irgendeiner Weise später verarbeitet werden können – er ermöglicht einfach ihre Einbettung in ein archivierbares Dokument.

Genau wie PDF/A-2 nicht PDF/A-1 ersetzt, ersetzt PDF/A-3 auch nicht PDF/A-2. Jeder Teil des PDF/A-Standards kann nach Bedarf zur Langzeitarchivierung verwendet werden.

Technische Konzepte von PDF/A

Grundlegende PDF/A-Anforderungen

PDF/A erfordert bestimmte PDF-Features and verbietet andere:

- ▶ Um die genaue visuelle Reproduzierbarkeit von Text zu gewährleisten, müssen alle in einem Dokument verwendeten Fonts eingebettet sein.
- ▶ Um die genaue Farb-Reproduzierbarkeit zu gewährleisten, müssen alle in einem Dokument verwendeten Farben in einer geräteunabhängigen Art und Weise definiert sein.
- ▶ Metadaten müssen im XMP-Format eingebettet sein. Die PDF/A-Konformitätsstufe muss mit bestimmten XMP-Properties beschrieben worden sein.
- ▶ Verschlüsselung darf nicht verwendet werden, um sicherzustellen, dass auf Dokumentinhalte ohne jede Einschränkung immer zugegriffen werden kann.
- ▶ Bestimmte Anforderungen an Anmerkungen und Formularfelder garantieren, dass die Visualisierung unverändert bleibt und dass Bildschirm- und Druckausgabe identisch sind.

Zusätzlich zu diesen einfachen Anforderungen erfordert PDF/A jedoch verschiedene andere PDF-Features (z.B. bestimmte Einträge in den Font-Datenstrukturen) und verbietet einige kritische Strukturen (z.B. bestimmte Kombinationen von TrueType-Fonts und -Encodings). Es gibt viele Aspekte, die umgesetzt und von Software-Entwicklern überprüft werden müssen, damit ein standardkonformes PDF/A-Produkt daraus entsteht, denn PDF/A ist viel mehr als einfach nur »PDF mit eingebetteten Fonts und ohne Verschlüsselung«!

Weitere Einschränkungen in PDF/A-1

PDF/A-1 leidet ein bisschen darunter, dass es der erste Standard in der PDF/A-Familie war: Die Norm wurde zu einer Zeit erstellt, als wichtige PDF-Konzepte noch nicht ausgereift waren. Als Ergebnis sind die folgenden Funktionen in PDF/A-1 verboten, aber in den neueren Teilen PDF/A-2 und PDF/A-3 erlaubt:

- ▶ Alle Funktionen, die PDF 1.5 oder höher erfordern, z.B. JPEG-2000-Kompression und Ebenen (optionale Inhalte).
- ▶ Transparenz: Obwohl Transparenz in PDF 1.4 möglich ist, wurde es für die Archivierung zu der Zeit als ungeeignet angesehen, weil es keine konsistente Beschreibung und Umsetzung für Transparenz-Unterstützung gab. Da identisches Verhalten nicht für alle PDF-Viewer garantiert werden konnte, wurde Transparenz in PDF/A-1 generell untersagt. Erst nach der Veröffentlichung von PDF/A-1 wurde die genaue Semantik der PDF-Transparenz geklärt und in ISO 32000-1 standardisiert; spätere Standards erlauben daher von vornherein die Verwendung von Transparenz.
- ▶ Dateianhänge waren in PDF/A-1 verboten, um alle Dokumentinhalte voll archivierbar zu machen. Obwohl Dateianhänge in PDF/A-2 erlaubt sind, werden sie hier auf PDF/A-1- oder PDF/A-2-konforme Dateien beschränkt, damit angehängte Dateien auf jeden Fall eindeutig reproduziert werden können. Bei PDF/A-3 werden diese Regeln weiter gelockert, hier sind beliebige Dateitypen als Anhang erlaubt.

Geräteunabhängige Farbspezifikation

Um eine konsistente Farbwiedergabe über Ausgabegeräte und Zeit hinweg sicherzustellen, verlangt PDF/A die Verwendung von geräteunabhängigen Farben. In der Regel wird dies über ICC-Profile oder CIE-Lab-Farbspezifikationen realisiert. Die optionale Ausgabebedingung beschreibt die Farbeigenschaften des Dokuments. Während diese Konzepte in der grafischen Industrie weit verbreitet sind, sind PDF-Entwickler in einem Unternehmen nicht unbedingt mit Farbmanagement vertraut und müssen sich erst mit ICC-Profilen und verwandten Konzepten beschäftigen.

Rasterbilder wie TIFF und JPEG spielen in der Dokumenterstellung eine entscheidende Rolle. Eingescannte Papierdokumente und Fotos aus Digitalkameras sind gängige Beispiele für Rasterbilddaten in Dokument-Workflows. In vielen Fällen sind Rasterbilddaten in modernen Arbeitsabläufen heute bereits geräteunabhängig, meist durch ein eingebettetes ICC-Farbprofil oder standardisierte Farbräume wie sRGB. Solche Bilder sind für den Einsatz in PDF/A bereits vorbereitet. Ältere Bilddaten sind dagegen in vielen Fällen geräteabhängig, wie Schwarzweiß- oder RGB-Scans ohne zugehöriges ICC-Profil.

XMP-Metadaten und Extension-Schemas

Extensible Metadata Platform (XMP) ist ein XML-basiertes Format, das auf Basis von RDF (*Resource Description Framework*) modelliert wurde, der Grundlage der semantischen Web-Initiative des W3C. Im Jahr 2012 wurde XMP als ISO 16684-1 standardisiert. PDF/A verlangt für die Speicherung von Informationen über ein Dokument die Verwendung von XMP-Metadaten innerhalb der PDF-Datei. XMP bietet ein leistungsstarkes und flexibles Framework für die Speicherung von vor- und benutzerdefinierten Eigenschaften für Metadaten.

Die XMP-Spezifikation umfasst mehr als ein Dutzend vordefinierte Schemas mit Hunderten sogenannter Properties für gängige Dokument- und Bildeigenschaften. Das am weitesten verbreitete vordefinierte XMP-Schema heißt Dublin Core. Es beinhaltet Eigenschaften wie Titel, Autor, Thema und Beschreibung.

XMP ist seinem Wesen nach erweiterbar, d.h. unternehmens- oder branchenspezifische Anforderungen an Metadaten können durch die Erstellung benutzerdefinierter Schemas erfüllt werden. PDF/A unterstützt dieses Konzept. Für eine automatisierte Abfrage verlangt PDF/A jedoch, dass zusätzlich eine maschinenlesbare Beschreibung der benutzerdefinierten Metadaten in das Dokument eingebettet ist. Dies wird durch eine »XMP-Extension-Schema-Beschreibung« erreicht: ein standardisierter Teil der XMP-Metadaten beschreibt den Aufbau der benutzerdefinierten Properties der XMP-Metadaten.

Konformität zu Level A: Tagged PDF

PDF/A-1a, PDF/A-2a und PDF/A-3a erfordern Tagged PDF. Während einfaches PDF lediglich den sichtbaren Inhalt auf einer Seite wiedergibt, muss für Tagged PDF die logische Dokumentstruktur innerhalb der Struktur-Hierarchie gespeichert werden. Tagged PDF bietet vordefinierte Strukturelement-Typen für die üblichen Teile eines Dokuments wie Überschriften, Tabellen und Listen. Sogenannte markierte Inhaltselemente können als Äquivalent des getaggtten Inhalts in Markup-Sprachen betrachtet werden. Sie beziehen sich auf Elemente in diesem Strukturbaum. Ähnlich wie HTML und XML unterstützt Tagged PDF Attribute für Strukturelemente. Zum Beispiel können Tabellenelemente Attribute für die Eigenschaften von Zeilen- oder Spaltenbreite einer Tabellenzelle tragen.

Die Konformität zu Level A erfordert auch, dass für den gesamten Text im Dokument Unicode-Semantik zur Verfügung steht und dass logische Wörter durch Leerzeichen getrennt sind.

PDF/UA-1 (Universal Accessibility, Barrierefreiheit oder Zugänglichkeit) ist ein neuer Standard, der viele Aspekte von Tagged PDF klärt. Er wurde im Jahr 2012 als ISO 14289 veröffentlicht. Obwohl es keine direkte Beziehung zwischen beiden Normen gibt, kann ein PDF/A-Dokument gleichzeitig konform zu einem PDF/UA-Dokument sein. Wir empfehlen sogar, die PDF/UA-Anforderungen einzuhalten, wenn Sie PDF/A mit Konformitätsstufe A erstellen möchten, um die Zugänglichkeit zu verbessern. Weitere Informationen hierzu können Sie unserem Whitepaper zu PDF/UA entnehmen.

Die folgende Einschränkung gilt für kombinierte PDF/A- und PDF/UA-Dokumente: Da PDF/UA das in PDF 1.4 und damit in PDF/A-1 fehlende Tabellenattribut *Scope* verlangt, können Tabellen mit Kopfzeilen nicht ordnungsgemäß getaggt werden. Bei der Verwendung von Tabellen empfehlen wir daher, statt PDF/A-1a die neueren Standards PDF/A-2a oder PDF/A-3a zu verwenden.

Konformität zu Level U: Unicode-Anforderungen

PDF/A-2 und PDF/A-3 bieten neben der Konformität zu Level A und B auch die Konformität zu Level U. Level U verlangt korrekte Unicode-Semantik für den gesamten Text im Dokument, jedoch kein Tagged PDF. Diese Anforderung rührt daher, dass PDF eine Vielzahl von Font- und Encoding-Techniken unterstützt, die nicht alle Unicode-fähig sind. Zum Beispiel unterstützt PDF die in den 1980er Jahren eingeführten PostScript-Type-1-Fonts, während das Unicode-Konsortium seine Arbeit erst im Jahr 1991 aufnahm. Für Level A und U müssen zusätzliche Unicode-Werte für alle Zeichen von solchen Fonts

enthalten sein, die diese Angaben nicht schon intern enthalten. Aber nicht alle Unicode-Werte sind akzeptabel: Werte in der Private Use Area (PUA) sind nicht erlaubt, da sie keine allgemeingültige Interpretation (Semantik) haben.

Symbolfonts wie Fonts mit Logos oder Piktogrammen sind ein wichtiger Anwendungsbereich dieser PDF/A-Anforderung. Da standardisierte Unicode-Werte für benutzerdefinierte symbolische Glyphen nicht verfügbar sind, muss geeignete Unicode-Semantik in einem als »ActualText« markierten Inhaltssattribut für den Text übergeben werden. Obwohl dieses Attribut üblicherweise nur in Tagged PDF verwendet wird, kann es auch in Dokumenten ohne Tags hinterlegt werden – genau das verlangt die Konformität zu Level U. »ActualText« kann einer einzelnen Glyphe oder einer Folge von mehreren Glyphen zugeordnet werden. Es kann aus einer beliebigen Unicode-Zeichenfolge bestehen.

Als Beispiel nehmen wir Code 0x1A aus dem gängigen Font Wingdings. Dieses Zeichen zeigt das Bild einer Computer-Tastatur mit dem Glyphnamen *keyboard* und dem PUA-Unicode-Wert U+F037. Der Glyphname könnte für die Erzeugung eines passenden *ActualText* verwendet werden, so zum Beispiel »Symbol für Keyboard«. Anzumerken ist, dass das programmatische Erzeugen von ActualText eine Notlösung bleibt; von einem menschlichen Bearbeiter ausgewählte Texte sind maschinengeneriertem ActualText immer vorzuziehen.

Konforme PDF/A-Viewer

Zwar sind alle PDF/A-konformen Dokumente auch PDF-Dokumente, nicht jedoch alle PDF-Viewer zwangsläufig PDF/A-koform. Dies liegt an den zusätzlichen Anforderungen, die vom PDF/A-Standard an PDF-Viewer gestellt werden. Das Konzept eines in der Norm definierten »PDF-Readers« betrifft nicht nur Tools zum interaktiven Anzeigen des Dokumentinhalts, sondern umfasst auch nicht-interaktive Tools wie einen Raster Image Processor (RIP). Während der grundlegende PDF-Prozess zur Darstellung eines Dokuments auf Bildschirm oder Papier in der ISO-Norm 32000-1 spezifiziert ist, schränkt PDF/A einige Aspekte der Darstellung weiter ein, wie zum Beispiel:

- ▶ Einfache PDF-Viewer dürfen ICC-basierte Farbspezifikationen ignorieren und stattdessen den in der PDF-Datei angegebenen alternativen Farbraum verwenden. PDF/A-konforme Reader müssen dagegen immer die geräteunabhängigen ICC-basierten Farbinformationen verwenden.
- ▶ PDF/A-konforme Reader müssen bestimmte geräteabhängige Informationen in einem Dokument wie etwa Schwarzaufbau und Unterfarbenreduktion (Undercolor Removal) ignorieren. (Das sind geräteabhängige Features für die Druckindustrie.)
- ▶ PDF/A-konforme Reader dürfen Dokumente nicht mit den lokal auf dem System installierten Fonts darstellen. Nur die im Dokument eingebetteten Fonts sind zur Darstellung erlaubt.
- ▶ Ab PDF/A-2 dürfen konforme Viewer nur noch auf XMP-Metadaten zurückgreifen statt auf herkömmliche Dokumentinfo-Felder.

Verarbeitung von PDF/A-Dokumenten

Besondere Sorgfalt muss auf die Verarbeitung von PDF/A-Dokumenten verwendet werden, um die Konformität mit dem Standard jederzeit zu erhalten. Selbst einfache Operationen können den Konformität eines Dokuments verletzen. Der Einsatz PDF/A-fähiger Tools ist deshalb von entscheidender Bedeutung, um zu vermeiden, dass Veränderungen an den Dokumenten zu einer Verletzung des Standards führen.

PDF-Dokumente aufteilen und verbinden

Selbst einfache Operationen können Standardverletzungen zur Folge haben. Bereits das Einfügen einer Seite in ein PDF/A-Dokument birgt zum Beispiel einige unmittelbare Gefahren:

- ▶ Wenn die eingefügte Seite aus einem nicht PDF/A-konformen Dokument stammt, könnte sie nicht eingebettete Fonts verwenden.
- ▶ Auch wenn die importierte Seite aus einem PDF/A-Dokument stammt, können in mehreren Bereichen Gefahren lauern. Zum Beispiel könnten die Farbeigenschaften (z.B. die Ausgabebedingung) der beiden Dokumente unterschiedlich sein, was wiederum zu einer nicht konformen Ausgabe führen könnte.
- ▶ Eine kleine Operation wie das Hinzufügen eines Metadaten-Feldes kann den Standard verletzen, sofern die Software die PDF/A-Regeln für XMP-Metadaten nicht richtig implementiert.

Jegliche Verarbeitung von Inhalten oder Metadaten eines PDF/A-Dokuments muss mit PDF/A-fähiger Software vorgenommen werden, um die PDF/A-Konformität zu erhalten.

Digitale Signaturen

Mit digitalen Signaturen in PDF-Dokumenten lässt sich die Integrität des Dokuments überprüfen, die Person, die die Signatur erstellt hat, authentisieren sowie Datum und Uhrzeit der Unterschrift fälschungssicher bestimmen. Integrierte digitale Signaturen sind Teil von PDF 1.4 und in PDF/A erlaubt. Mehrfach-Signaturen sind mit der Ausnutzung der inkrementellen Update-Funktion von PDF ebenfalls erlaubt. Allerdings müssen die Signaturen für PDF/A bestimmten Anforderungen genügen:

- ▶ Wenn die Signatur ein visuelles Erscheinungsbild hat (z.B. ein Bild oder eine textuelle Darstellung der Namen des Unterzeichners), muss dieses den gleichen PDF/A-Anforderungen entsprechen wie andere Teile des Dokuments (geräteunabhängige Farbe, eingebettete Fonts usw.).
- ▶ PDF/A-2 enthält zusätzliche Anforderungen in Bezug auf die technischen Details der Signatur. Die Norm empfiehlt, Informationen zu Zeitstempeln und Zertifikatssperlisten in die Signatur mit aufzunehmen.
- ▶ Der Signatur-Prozess sollte in den XMP-Metadaten des Dokuments vermerkt werden.

Um den Einsatz von digitalen Signaturen in PDF/A-Workflows zu nutzen, muss das Signatur-Tool PDF/A-fähig sein, das heißt die oben skizzierten Regeln einhalten.

Kurz gesagt sollten nur PDF/A-fähige Tools in PDF/A-Workflows verwendet werden, andernfalls kann die PDF/A-Konformität leiden. Um PDF/A-Verstöße durch versehentliche Änderung zu vermeiden, öffnet Adobe Acrobat PDF/A-Dokumente standardmäßig im Lesemodus (read-only).

PDF/A-Unterstützung in der PDFlib-Produktfamilie

PDFlib GmbH führte bereits 2006 die PDF/A-Funktionalität in ihren Produkten ein. Die Produkte von PDFlib waren die ersten, die XMP-Extension-Schemas unterstützen. Alle Produkte der aktuellen Version 9 der Produktfamilie PDFlib unterstützen alle Varianten von PDF/A-1, PDF/A-2 und PDF/A-3. Anwendungsentwicklern wird ein Toolkit für die folgenden für PDF/A relevanten Operationen an die Hand gegeben:

- ▶ PDF/A von Grund auf neu erzeugen, z.B. basierend auf Text aus einer Datenbank
- ▶ Rasterbilder nach PDF/A konvertieren (z.B. Scans)
- ▶ bestehende PDF/A-Dokumente bearbeiten, z.B. aufteilen oder zusammensetzen
- ▶ Verwendung von ICC-Profilen und geräteunabhängiger Farbe für alle Bereiche des Color-Managements
- ▶ Erzeugung von PDF/A Level A mit Strukturinformation (Tagged PDF), auch in Kombination mit PDF/UA
- ▶ das generierte Dokument mit XMP-Metadaten versehen, einschließlich XMP-Extension-Schemas
- ▶ Bestimmung von Glyphnamen in Fonts zur Bestimmung eines geeigneten ActualText für Symbol-Fonts
- ▶ Anhängen von PDF/A-Dokumenten an PDF/A-2-Dateien oder von beliebigen Dateitypen an PDF/A-3-Dateien

Diese Operationen können mit einfachen PDFlib-Funktionsaufrufen implementiert werden. Mit den PDFlib-Paketen wird Beispielcode für eine Vielzahl an Programmiersprachen und Entwicklungsumgebungen ausgeliefert. Im PDFlib-Cookbook werden weitere Programmiermethoden vorgestellt. Um die von PDF/A geforderte Font-Einbettung zu erleichtern, enthält das japanische Resource-Kit von PDFlib gängige japanische Fonts. In der Software-Lizenz ist bereits eine Lizenz für die Einbettung dieser Fonts enthalten.

Erzeugung von PDF/A mit PDFlib

Mit PDFlib kann auf Basis folgender Mechanismen PDF/A-konforme Ausgabe erzeugt werden:

- ▶ PDFlib kümmert sich automatisch um verschiedene formale Einstellungen für PDF/A, zum Beispiel die PDF-Version und die erforderlichen XMP-Identifikationseinträge.
- ▶ Der PDFlib-Client muss explizit bestimmte Funktionsaufrufe und Optionen verwenden (z.B. für die Font-Einbettung).
- ▶ Der PDFlib-Client darf bestimmte Funktionsaufrufe und Optionen nicht verwenden (z.B. Verschlüsselung).

Wenn sich der PDFlib-Client an diese Regeln hält, ist standardkonforme PDF/A-Ausgabe gewährleistet. Entdeckt PDFlib eine Regelverletzung bei der PDF/A-Erstellung, so wird eine Exception ausgelöst, die von der Anwendung abgefangen werden muss.

In diesem Fall wird keine PDF-Ausgabe erzeugt. Die Gefahr der Erzeugung nichtkonformer Ausgabe beim Auftreten eines Fehlers besteht nicht. Details erforderlicher und verbotener Operationen werden in der PDFlib-Dokumentation erläutert.

Verarbeitung von PDF/A mit PDFlib

Beim Importieren von Seiten aus vorhandenen PDF/A-Dokumenten müssen weitere Regeln beachtet werden. Beim Umgang mit bestehenden PDF/A-Dokumenten prüft PDFlib + PDI sorgfältig die PDF/A-Eigenschaften aller Ein- und Ausgabedokumente, um sicherzustellen, dass die Ausgabe noch PDF/A-konform ist. Als zusätzliche Steuermöglichkeit kann die Ausgabebedingung eines importierten Dokuments in die Ausgabe-PDF kopiert werden, um die PDF/A-Farbeeigenschaften eines vorhandenen Dokuments exakt zu reproduzieren. Ebenso können XMP-Metadaten aus importierten Dokumenten kopiert oder zusammengefasst werden.

Erzeugung von PDF/A Level A mit PDFlib

PDF/A Level A kann als Level B mit Tagged PDF betrachtet werden. Die Unterstützung von PDFlib für PDF/A Level A basiert auf den Funktionen zur Erzeugung von Tagged PDF: Jedes Inhaltselement kann an einer bestimmten Stelle im Strukturbaum des Dokuments platziert werden. Für die Dokumentstruktur nicht relevante Elemente wie Kopf-, Fußzeilen und Seitenzahlen können als Artefakte ausgezeichnet werden. Damit werden sie ignoriert, wenn das Dokument von einer Software vorgelesen oder in ein anderes Format konvertiert wird. Für Rasterbilder und Vektorgrafik kann Alternativtext hinterlegt werden. PDFlib zeichnet Tabellen und Artefakte automatisch aus, was eine große Zeitersparnis für Entwickler bedeutet und die Erstellung standardkonformer Strukturhierarchie deutlich vereinfacht. PDFlib prüft die übergebenen Tags, um sicherzustellen, dass die Verschachtelung der Strukturelemente und -Attribute ISO 32000-1-konform ist. Zum Beispiel müssen Tags für Überschriften oder Listen korrekt verschachtelt sein.

Die integrierte Unterstützung von PDF/UA erleichtert die Erzeugung von sowohl zugänglicher als auch archivierbarer PDF-Ausgabe. Beachten Sie, dass zur Erzeugung von Tagged PDF detaillierte Kenntnisse der logischen Dokumentstruktur benötigt werden. PDFlib kümmert sich zwar um die PDF-bezogenen Informationen, aber es kann nicht die Dokumentstruktur aus dem Inhalt des Dokuments ableiten.



PDFlib GmbH

Franziska-Bilek-Weg 9
80339 München, Germany
phone +49 • 89 • 452 33 84-0
support@pdflib.com
www.pdflib.com/knowledge-base/pdfa

PDFlib GmbH ist auf die Entwicklung von PDF-Technologie spezialisiert. PDFlib-Produkte sind seit 1997 weltweit im Einsatz. Das Unternehmen berücksichtigt wichtige technologische Trends, etwa ISO-Standards für PDF. PDFlib GmbH vertreibt alle Produkte weltweit, wobei Nordamerika, Europa und Japan die wichtigsten Märkte darstellen.

