



## PDFlib pCOS 3

PDF Information Retrieval Tool

### Was ist PDFlib pCOS?

PDFlib pCOS bietet eine einfache und elegante Methode, um aus PDF-Dokumenten Informationen abzurufen, die nicht zum Seiteninhalt gehören. PDF-Metadaten, interaktive Elemente (Links etc.) oder Seitengrößen sind zum Beispiel bequem mit pCOS abfragbar.

Mit pCOS können Sie zahlreiche interessante Elemente auslesen und für verschiedene Verwendungszwecke ausgeben. Da pCOS in einem einzigen Aufruf viele PDF-Dokumente verarbeiten kann, können Sie sehr einfach Übersichten über Dokumentinfelder, Seitengrößen, Schriften und andere Eigenschaften erstellen. In Kombination mit tabellarischer Ausgabe erwächst daraus ein leistungsfähiges PDF-Verwaltungswerkzeug.

Für PDFlib pCOS (PDF Information Retrieval Tool) gibt es vielseitige Einsatzmöglichkeiten in PDF-Workflows. Sie können PDFlib pCOS aber auch zur PDF-Fehlersuche oder zum Erlernen des Dateiformats verwenden. Einige typische Anwendungsfälle:

- ▶ Eingehende Dokumente auf bestimmte Eigenschaften prüfen
- ▶ Problemfälle in umfangreichen Dokumentbeständen ermitteln
- ▶ Metadaten für Dokumentenmanagement zusammenfassen
- ▶ Qualitätskontrolle von Dokumenten vor der Veröffentlichung
- ▶ Workflows zur Speicherung und Bereitstellung von Dokumenten
- ▶ Lesezeichen zusammenfassen
- ▶ Komponenten von PDF-Dokumenten extrahieren, z.B. ICC-Profile
- ▶ PDFs auf Sicherheitsrisiken überprüfen (JavaScript etc.)

Die pCOS-Schnittstelle ist auch in weiteren Produkten von PDFlib GmbH enthalten: wenn Sie PDFlib+PDI, PDFlib Personalization Server, TET oder PLOP benutzen, können Sie die pCOS-Schnittstelle ebenfalls nutzen. Um Text oder Rasterbilder aus PDF-Seiten zu extrahieren nutzen Sie bitte unser Produkt PDFlib TET.

#### pCOS-Cookbook

Das pCOS-Cookbook ist eine Sammlung von Programmierbeispielen, die zeigen, wie sich pCOS für verschiedene Anwendungsfälle zur Informationsbeschaffung nutzen lässt. Das Cookbook steht auf der PDFlib-Website zur Verfügung und enthält Beispiel-Code, Eingabedokumente und exemplarische pCOS-Ausgabe.

### Funktionalität von PDFlib pCOS

#### Unterstützte Eingabeformate

PDFlib pCOS verarbeitet alle Varianten von PDF:

- ▶ Alle PDF-Versionen bis Acrobat X einschließlich ISO 32000
- ▶ Verschlüsselte Dokumente (ggf. ist das Kennwort erforderlich)
- ▶ Ausgefeiltes Sicherheitsmodell: selbst ohne Kennwort erhalten Sie bestimmte Informationen, solange dies nicht die Interessen des Dokumentverfassers verletzt
- ▶ Beschädigte PDF-Eingabedokumente werden repariert

#### Abfragen von Informationen aus PDF-Dokumenten

PDFlib pCOS bietet eine einfach zu nutzende Abfrageschnittstelle. Mit pCOS können Sie zahlreiche interessante Elemente auslesen, zum Beispiel:

- ▶ Dokumentinfelder und XMP-Metadaten
- ▶ Allgemeine Angaben: Status von Linearisierung und Tagged PDF, Verschlüsselungsinformationen und Berechtigungseinstellungen, Anzahl von Seiten und Schriften
- ▶ Schriften mit Name, Einbettungsstatus, etc.
- ▶ Rasterbildinformationen: Größe, Bittiefe, Farbraum, Kompression
- ▶ Details zu Farbräumen
- ▶ Verweisziele (URLs) und Koordinaten von Web-Links
- ▶ Lesezeichen mit Seitennummern, um ein Inhaltsverzeichnis zu erstellen
- ▶ Formularfelder: vollständiger Feldname, Inhalt, Position, etc.
- ▶ Seitengröße, CropBox, Seitendrehung
- ▶ Status von ISO-Standards: PDF/X, PDF/A, PDF/UA, PDF/E, PDF/VT
- ▶ Geo-Referenzdaten
- ▶ Dateianhänge auflisten oder extrahieren
- ▶ Namen von Ebenen, symbolische Seitennamen, Artikel
- ▶ Details zu Anmerkungen
- ▶ Alle Kommentare mit Namen des jeweiligen Bearbeiters
- ▶ Details zu digitalen Signaturen: Namen der Signaturfelder, signiert/unsigned, Name des Unterzeichners, Datum und Grund der Signatur
- ▶ ICC-Profile für Druckausgabebedingungen aus PDF/X- oder PDF/A-Dateien extrahieren
- ▶ Eigenschaften von Blöcken für PDFlib Personalization Server
- ▶ JavaScript auf Dokument-, Seiten-, Anmerkungs- und Feldebene

## Ausgabeformate

PDFlib pCOS erzeugt Ausgabe für verschiedene Zwecke:

- ▶ Einfacher Text
- ▶ Unicode-Text in den Formaten UTF-8 und UTF-16
- ▶ Tabellarische Ausgabe zur Weiterverarbeitung in einer Tabellenkalkulation oder Datenbank
- ▶ Extraktion von Binärdaten, z.B. ICC-Profile oder Dateianhänge
- ▶ Benutzerdefinierte Formate zur individuellen Weiterverarbeitung

## pCOS-Pfade: einfache Syntax für PDF-Objekte

Mit der übersichtlichen pCOS-Pfadsyntax können Sie auf PDF-Objekte zugreifen, ohne sich in komplizierten Baumstrukturen, wie bei Lesezeichen oder Formularfeldern, zu verzetteln. Die pCOS-Pfadsyntax bietet praktische Abkürzungen, mit denen Sie auf häufig benötigte PDF-Objekte zugreifen können, etwa Seiten, Schriften, Lesezeichen oder Formularfelder.

## pCOS als Bibliothek oder Kommandozeilen-Tool?

pCOS wird als Software-Bibliothek (Komponente) für verschiedene Entwicklungsumgebungen sowie als Kommandozeilen-Tool für Batch-Prozesse ausgeliefert. Beide Ausführungen bieten den gleichen Funktionsumfang, eignen sich aber für unterschiedliche Einsatzbereiche.

## Die pCOS-Software-Bibliothek eignet sich...

...zur Integration in Desktop- oder Server-Anwendungen. Programmierbeispiele für alle unterstützten Sprachbindungen sind im pCOS-Paket enthalten.

## Das pCOS-Kommandozeilen-Tool eignet sich...

...zur Batch-Verarbeitung von PDF-Dokumenten. Es erfordert keine Programmierung, sondern kann über leistungsfähige Kommandozeilen-Optionen gesteuert und damit in komplexe Arbeitsabläufe integriert werden. Das pCOS-Kommandozeilen-Tool ergänzt die Funktionen der Bibliothek um:

- ▶ Einfache Abfrage typischer PDF-Elemente, z.B. Lesezeichen, Anmerkungen, Metadaten, Formularfelder, etc.
- ▶ Erweiterte Abfrage komplexer Objekte und Steuerung des Ausgabeformats
- ▶ Elemente wie Dateianhänge, ICC-Profile etc. extrahieren
- ▶ Ausgabe der Informationen als CSV oder in benutzerdefiniertem Format zum Import in Tabellenkalkulation oder Datenbank
- ▶ Rekursion zur Erfassung verschachtelter Objekte, z.B. Dictionaries oder Arrays

## Unterstützte Entwicklungsumgebungen

PDFlib pCOS läuft überall – auf praktisch allen Computersystemen. Wir bieten 32- und 64-Bit-Pakete an und unterstützen alle gängigen Varianten von Windows, Mac OS X, Linux und Unix.

Der Kern von pCOS ist in C und C++ geschrieben und auf Schnelligkeit und geringen Overhead optimiert. Über ein einfaches API (Application Programming Interface) lässt sich die pCOS-Funktionalität in zahlreichen Programmiersprachen nutzen:

- ▶ COM für VB, ASP und zahlreichen anderen Sprachen
- ▶ C und C++
- ▶ Java einschließlich Servlets und Java Application Server
- ▶ .NET für C#, VB.NET, ASP.NET, etc.
- ▶ Perl
- ▶ PHP
- ▶ Python

## Vorteile von PDFlib-Software

### Zuverlässig

Weltweit arbeiten viele Tausend Programmierer erfolgreich mit unserer Software. PDFlib-Produkte erfüllen alle Qualitäts- und Geschwindigkeitskriterien für den Einsatz auf großen Servern. Alle Produkte sind für den zuverlässigen, unbeaufsichtigten 24-Stunden-Betrieb ausgelegt.

### Schnell und einfach

PDFlib-Produkte sind unglaublich schnell – bis zu Tausenden von Seiten pro Sekunde. Die Programmierschnittstelle ist übersichtlich und einfach zu erlernen.

### PDFlib-Produkte sind überall

Unsere Produkte unterstützen alle internationalen Sprachen sowie Unicode. Sie werden von Kunden in der ganzen Welt eingesetzt.

### Professioneller Support

Bei Problemen bietet Ihnen unser Support-Team professionelle Unterstützung. Um den reibungslosen Ablauf unternehmenskritischer Anwendungen zu gewährleisten, können Sie Ihre Software-Lizenz durch einen Supportvertrag ergänzen. Ein Supportvertrag garantiert Ihnen kurze Antwortzeiten und Zugang zu den jeweils neuesten Versionen.

### Lizenzierung

Bei der Lizenzierung können Sie zwischen verschiedenen Modellen für Server-, Integrations-, Firmen- sowie Quellcodelizenzen wählen. Ergänzend bieten wir Supportverträge für umfangreichen technischen Support mit kurzen Reaktionszeiten und kostenlosen Software-Aktualisierungen an.

## Über PDFlib GmbH

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.

## Kontakt

Evaluierungsversionen mit vollem Funktionsumfang und Dokumentation sowie Beispielen sind auf unserer Webseite verfügbar. Weitere Informationen erhalten Sie unter:



### PDFlib GmbH

Franziska-Bilek-Weg 9, 80339 München

Tel. +49 • 89 • 452 33 84-0, Fax +49 • 89 • 452 33 84-99

sales@pdfliib.com

www.pdfliib.com