



Datenblatt

pCOS

*Schnittstelle zum
Abfragen von
Informationen
aus PDF*

Was ist die pCOS-Schnittstelle?

pCOS bietet eine einfache und elegante Methode, um aus PDF-Dokumenten Informationen abzurufen, die nicht zum Seiteninhalt gehören (Seiteninhalte können mit PDFlib TET extrahiert werden). Die pCOS-Schnittstelle ist kein eigenständiges Produkt, sondern integrierter Bestandteil der folgenden Produkte:

- ▶ PDFlib+PDI und PDFlib Personalization Server (PPS) (pCOS ist nicht im Basisprodukt PDFlib enthalten)
- ▶ Text and Image Extraction Toolkit (PDFlib TET)
- ▶ In TET PDF IFilter können Sie mit pCOS Informationen aus PDF-Dokumenten abfragen und zur Indizierung und Suche benutzen.
- ▶ PDFlib PLOP/PLOP DS

Diese Produkte enthalten die pCOS-Programmierschnittstelle. PLOP/PLOP DS-Pakete enthalten zusätzlich das pCOS-Kommandozeilenwerkzeug.

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 Dokument-Infelder, Seitengrößen, Schriften und andere Eigenschaften erstellen. In Kombination mit tabellarischer Ausgabe erwächst daraus ein leistungsfähiges PDF-Verwaltungswerkzeug.

Für pCOS gibt es vielseitige Einsatzmöglichkeiten in PDF-Workflows. Sie können pCOS aber auch zur PDF-Fehlersuche oder zum Erlernen des Dateiformats verwenden, z.B.:

- ▶ 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.)

pCOS-Cookbook

Das pCOS-Cookbook ist eine Sammlung von Programmierbeispielen, die zeigen, wie sich pCOS zum Auslesen von PDF-Informationen nutzen lässt. Das Cookbook steht auf der PDFlib-Website zur Verfügung und enthält Code, Eingabedokumente und pCOS-Ausgabe.

Funktionalität von pCOS

Abfragen von Informationen aus PDF-Dokumenten

Mit pCOS können Sie zahlreiche interessante Elemente auslesen, zum Beispiel:

- ▶ Dokument-Infelder und XMP-Metadaten
- ▶ allgemeine Angaben: Status von Linearisierung und Tagged PDF, Verschlüsselungsinformationen und Berechtigungseinstellungen, Anzahl von Seiten und Schriften
- ▶ Schriften mit Name, Einbettungsstatus, etc.
- ▶ Rasterbilder: Größe, Bittiefe, Farbraum, Kompression, XMP
- ▶ 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/A, PDF/E, PDF/UA, PDF/VCR, PDF/VT und PDF/X
- ▶ Geo-Referenzdaten
- ▶ Dateianhänge auflisten oder extrahieren
- ▶ Namen von Ebenen, symbolische Seitennamen, Artikel
- ▶ Details zu Anmerkungen
- ▶ Kommentare mit Namen des jeweiligen Bearbeiters
- ▶ Details zu digitalen Signaturen: Namen der Signaturfelder, signiert/unsigned, Name des Unterzeichners, PAdES
- ▶ 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
- ▶ XML-Rechnungsdaten aus ZUGFeRD-Dokumenten extrahieren
- ▶ Eigenschaften von PDF-Paketen/-Portfolios

Unterstützte Eingabeformate

pCOS verarbeitet alle Varianten von PDF:

- ▶ Alle PDF-Versionen bis zu Acrobat DC, also PDF 1.7 (ISO 32000-1) einschließlich Extension Level 8 sowie PDF 2.0 (ISO 32000-2)
- ▶ Verschlüsselte Dokumente (ggf. ist das Kennwort erforderlich)
- ▶ Beschädigte PDF-Eingabedokumente werden repariert

Ausgabeformate

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-Profilen oder Dateianhängen
- ▶ 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 Kommandozeilenwerkzeug?

pCOS ist als Software-Bibliothek (Komponente) für verschiedene Entwicklungsumgebungen sowie als Kommandozeilenwerkzeug für Batch-Verarbeitung verfügbar. Beachten Sie, dass das Kommandozeilenwerkzeug nur in den Produktpaketen von PDFlib PLOP und PLOP DS enthalten ist. Beide Ausführungen bieten den gleichen Funktionsumfang, eignen sich aber für unterschiedliche Einsatzbereiche.

Die pCOS-Programmierschnittstelle eignet sich...

...zur Integration in Desktop- oder Server-Anwendungen. Programmierbeispiele für alle unterstützten Sprachbindungen sind in den Produktpaketen enthalten. Viele weitere Beispiele finden Sie im pCOS-Cookbook.

Das pCOS-Kommandozeilenwerkzeug 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-Kommandozeilenwerkzeug 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-Profilen etc. extrahieren
- ▶ Ausgabe der Informationen als CSV oder in benutzerdefiniertem Format zum Import in Tabellenkalkulation oder Datenbank
- ▶ Rekursive Ausgabe verschachtelter Objekte, z.B. Dictionaries oder Arrays

Evaluierung

Alle Produkte, die die pCOS-Schnittstelle enthalten, sind zur Evaluierung verfügbar. Voll funktionsfähige Evaluierungsversionen mit Dokumentation und Beispielen sind auf unserer Webseite verfügbar.

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. Unsere Produkte sind seit 1997 im Einsatz. 2006 waren wir eines der Gründungsmitglieder der PDF Association (damals noch PDF/A Competence Center). Das Unternehmen berücksichtigt wichtige technologische Trends, etwa ISO-Standards für PDF. PDFlib GmbH vertreibt alle Produkte weltweit, wobei Europa, Nordamerika und Japan die wichtigsten Märkte darstellen.

Kontakt

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



PDFlib GmbH

Franziska-Bilek-Weg 9, D-80339 München

Tel. +49 · 89 · 452 33 84-0

sales@pdflib.com www.pdflib.com