



## pCOS

PDF情報取得  
インタフェース

## pCOS インタフェースとは

pCOS は、PDF 文書からページ内容以外の任意の情報を取得するための、シンプルかつエレガントな機構を提供します（ページ内容については PDFlib TET を用いて抽出できます）。pCOS インタフェースは、独立の製品ではなく、以下の製品に組み込まれています：

- ▶ PDFlib+PDI・PDFlib Personalization Server (PPS) (ベース PDFlib 製品は pCOS を内蔵していません)
- ▶ テキスト・画像抽出ツールキット (PDFlib TET)
- ▶ PDFlib TET PDF IFilter では、pCOS を使って PDF 文書から情報を抽出し、それを利用してインデックス作成や検索が可能です。
- ▶ PDFlib PLOP・PLOP DS

以上の製品は pCOS プログラミングインタフェースを内蔵しています。PLOP・PLOP DS パッケージはさらに pCOS コマンドラインツールを内蔵しています。

pCOS を用いれば、さまざまな有用なアイテムを抽出することができ、多様な用途へ向けて出力を生成することが可能です。複数の PDF 文書をただ一回の呼び出しで処理して、文書情報項目・判型・フォント等、任意のプロパティの一覧を容易に生成できます。これに表形式出力を組み合わせれば、強力な PDF 管理ツールを実現可能です。

pCOS は、さまざまな応用シナリオの PDF ワークフローで利用できるのみならず、PDF を学習したりデバッグしたりするためのツールとしても有用です。典型的な活用シチュエーションは以下の通りです：

- ▶ 受領する文書が、あらかじめ定義しておいた条件に適合しているかを検査
- ▶ 大量の文書の中から、問題のあるファイルを発見
- ▶ メタデータのサマリを生成して文書管理に活用
- ▶ 文書を発行する前に品質を確認
- ▶ 各種の文書検索リポジトリのワークフロー
- ▶ ブックマークの一覧を生成
- ▶ ICC プロファイル等、PDF 文書の各種の構成要素を抽出
- ▶ PDF にセキュリティ上問題がないかを検査 (JavaScript 等)

## pCOS Cookbook

pCOS Cookbook は、さまざまな PDF 抽出のタスクに pCOS を活用する例を示すプログラミングサンプル集です。この Cookbook は PDFlib ウェブサイトにあり、コードと入力文書と pCOS 出力があります。

## pCOS のさまざまな機能

## 情報取得

pCOS を用いて、さまざまな有用なアイテムを抽出できます：

- ▶ 各種の文書情報項目と XMP メタデータ
- ▶ 一般的情報：PDF の線形化・タグ付きステータス、暗号化の詳細と権限設定、ページ数やフォント
- ▶ フォントの名称・埋め込みステータス等
- ▶ 画像のビット深度・色空間・圧縮・XMP 等のデータ
- ▶ 色空間の詳細
- ▶ Web リンクの移動先 URL や座標
- ▶ しおりとそのページ番号。目次の生成等に役立ちます。
- ▶ フォームフィールドデータ：フィールドの完全名・内容・位置等
- ▶ ページ寸法、CropBox やページ回転
- ▶ 各種 ISO 規格のステータス：PDF/A・PDF/E・PDF/UA・PDF/VCR・PDF/VT・PDF/X
- ▶ 地理空間参照情報
- ▶ ファイル添付のリストや抽出
- ▶ レイヤー名、ページラベルやアークスレッド
- ▶ 注釈の詳細
- ▶ コメントとそのレビューア名をリスト
- ▶ 電子署名の詳細：署名フィールドの名称・署名済/未署名・署名者名・PAdES
- ▶ ICC 出力インテントプロファイルを PDF/X・PDF/A 文書から抽出
- ▶ PDFlib Personalization Server のための各種プロックプロパティ
- ▶ 文書・ページ・注釈・フィールドレベルの JavaScript
- ▶ ZUGFeRD 文書から XML インボイスデータを取得
- ▶ PDF パッケージ/ポートフォリオの各種プロパティ

## 対応する入力

pCOS はあらゆる種類の PDF 入力に対応しています：

- ▶ Acrobat DC すなわち PDF 1.7 (ISO 32000-1) 拡張レベル 8 までのすべての PDF バージョンおよび PDF 2.0 (ISO 32000-2)
- ▶ 暗号化文書 (パスワードが必要な場合があります)
- ▶ PDF 入力文書が破損している場合には、可能な限り修復

## 出力形式

pCOS は、さまざまな目的に向けて出力を生成できます：

- ▶ プレーンテキスト出力
- ▶ UTF-8 または UTF-16 形式の Unicode テキスト出力
- ▶ スプレッドシートやデータベースで扱える表形式出力
- ▶ ICC プロファイルやファイル添付等のバイナリデータ
- ▶ カスタムな後処理のため、ユーザー定義の出力形式

## pCOS パス：各種 PDF オブジェクトへのシンプルな文法

しおりやフォームフィールド等へ至る複雑な木構造に疲弊する必要なく、シンプルな pCOS パス文法を使って簡単にさまざまな PDF オブジェクトにアクセスできます。ページ・フォント・しおり・フォームフィールド等、よく使われる PDF オブジェクトにアクセスする便利なショートカットとして役立ちます。

## pCOS ライブラリコマンドラインツールか

pCOS は、さまざまな開発環境用のプログラミングライブラリ (コンポーネント) としても、バッチ処理のためのコマンドラインツールとしても利用できます。なお、pCOS コマンドラインツールは、PDFlib PLOP と PLOP DS の製品パッケージにのみ内蔵されています。プログラミングインタフェースとコマンドラインツールは提供する機能は同様ですが適した運用タスクが異なります。

## pCOS プログラミングライブラリの用途は…

…デスクトップ/サーバアプリケーションへの組み込みです。このライブラリのすべての対応言語バインディングでのシンプルな使用例が各製品パッケージに同梱されています。さらに多くの例が pCOS Cookbook にあります。

## pCOS コマンドラインツールの得意分野は…

…PDF 文書のバッチ処理です。プログラミングを一切必要とせずに強力なコマンドラインオプション群を提供して複雑なワークフローへの統合を可能にします。pCOS コマンドラインツールは、ライブラリの機能を拡張します：

- ▶ しおり・注釈・メタデータ・フォームフィールド等、PDF でよく使われるさまざまな要素をシンプルに取得
- ▶ より複雑なオブジェクトをクエリしたり、出力形式をカスタマイズしたりすることが可能な拡張モード
- ▶ ファイル添付や ICC プロファイル等、各種データアイテムを抽出
- ▶ カンマ区切りやユーザーが定義した形式で情報を出力することにより、スプレッドシートやデータベースへの取り込みを可能に
- ▶ 辞書や配列等、複合的な PDF オブジェクトをダンプできる再帰機能

## 評価版

pCOS インタフェースを内蔵している製品にはすべて評価版があります。完全に機能する評価版が、説明書とサンプルとともに当社 Web サイトにあります。

## PDFlib ソフトウェア利用の利点

### 磐石の製品群

世界中の数万人のプログラマーが当社のソフトウェアを使用して成功しています。PDFlib 製品群はサーバ運用のためのあらゆる品質・パフォーマンス要求を満たします。製品はすべて、堅牢な無休サーバ運用と無人バッチ処理に適しています。

### 速くてシンプル

PDFlib 製品群は非常に高速です——秒速数千ページを実現します。そのプログラミングインタフェースは簡明で学習が容易です。

### 世界中に PDFlib 製品

当社製品群は世界のあらゆる言語と Unicode に対応しています。世界のあらゆる場所のお客様にご利用いただいています。

### プロフェッショナルサポート

問題があるとき、当社は支援に努めます。ビジネスクリティカルなさまざまな応用の要求を満たす商用サポートを提供しています。サポートを追加すると、最新バージョンへのアクセスと、万一の問題発生時の回答時間保証をご利用いただけます。

### ライセンスング

サーバライセンス・統合/サイトライセンス・ソースコードライセンスのためのさまざまなライセンスングオプションを提供しています。短時間回答と無償アップデートを伴う幅広い技術サポートのためのサポート契約もご利用いただけます。

## PDFlib GmbH について

PDFlib GmbH は PDF 技術に完全特化しています。世界中のお客様が PDFlib 製品群を 1997 年から利用しています。PDF 関連 ISO 規格群等の技術・市場動向に密に追随しています。PDFlib GmbH 製品群は世界中へ出荷されており、主要市場は北米・欧州・日本となっています。

## お問い合わせ

詳しくはお問い合わせください：



### PDFlib GmbH

Franziska-Bilek-Weg 9, 80339 München, Germany  
電話 +49・89・452 33 84-0、FAX +49・89・452 33 84-99  
sales@pdflib.com  
www.pdflib.com