



## Che cos'è PDFlib pCOS?

PDFlib pCOS offre un metodo semplice ed elegante per recuperare le informazioni, anche non visibili, contenute all'interno di un documento PDF. Con pCOS è possibile ad esempio leggere i metadati, gli elementi interattivi (link, campi, ecc.) o le dimensioni della pagina.

Attraverso pCOS potete recuperare numerosi elementi di interesse ottenendo degli output utili per diversi scopi. Grazie all'elaborazione contemporanea di numerosi documenti PDF è possibile creare in modo semplice rapporti contenenti informazioni sui campi, sui formati della pagina, sui font e su altre proprietà. Insieme al formato tabellare per fogli di calcolo, pCOS offre uno strumento di amministrazione particolarmente potente.

Le possibili applicazioni di PDFlib pCOS all'interno dell'Information Retrieval da PDF sono numerose, ma PDFlib pCOS può essere utilizzato anche come strumento per la ricerca degli errori nei file PDF o per approfondire le conoscenze sul formato PDF. Alcune tipiche applicazioni:

- ▶ Controllare determinate caratteristiche dei documenti in ingresso
- ▶ Identificare i file corrotti in grandi serie di documenti
- ▶ Creare indici di metadati per la gestione del documento
- ▶ Controllare la qualità prima della pubblicazione dei documenti
- ▶ Workflow per il recupero e la memorizzazione di documenti
- ▶ Creare rapporti sui segnalibri
- ▶ Estrarre componenti dai file PDF, ad esempio i profili ICC
- ▶ Controllare la presenza di problemi inerenti la sicurezza nei PDF (JavaScript, ecc.)

L'interfaccia per il recupero di informazioni di pCOS è integrata in altri prodotti PDFlib GmbH: utilizzando PDFlib+PDI, PDFlib Personalization Server, TET o PLOP si avrà accesso all'interfaccia di pCOS. Se si desidera accedere al testo o alle immagini, lo strumento ideale è PDFlib TET per l'estrazione dei contenuti da PDF.

### pCOS Cookbook

pCOS Cookbook è una raccolta di programmi dimostrativi che illustrano le varie possibilità di impiego di pCOS per l'estrazione di testo e immagini. Il Cookbook, disponibile nel sito di PDFlib,

## Datasheet

# PDFlib pCOS 3

## PDF Information Retrieval Tool

contiene script dimostrativi, documenti di prova ed esempi di risultati ottenibili.

## Caratteristiche di PDFlib pCOS

### Input supportati

PDFlib pCOS supporta tutte le principali varianti di PDF:

- ▶ Tutte le versioni PDF fino ad Acrobat X, compreso ISO 32000
- ▶ Documenti crittografati (eventualmente con richiesta password)
- ▶ Sistemi di sicurezza avanzati: anche senza avere a disposizione la password, è possibile ottenere diverse informazioni se queste non violano gli interessi dell'autore del documento
- ▶ I PDF danneggiati, se possibile, vengono riparati

### Information Retrieval

PDFlib pCOS offre una semplice interfaccia di interrogazione. Con PDFlib pCOS è possibile recuperare numerose informazioni:

- ▶ Informazioni sul documento e metadati XMP
- ▶ Informazioni generali: stato di linearizzazione, etichette, crittografia, permessi, numero di pagine e font
- ▶ Tutti i font allegati con nomi, stato, ecc.
- ▶ Dati sulle immagini, ad es. profondità di bit, spazio colore, compressione
- ▶ Dettagli sullo spazio colore
- ▶ URL target e coordinate dei link web
- ▶ Tutti i bookmark con corrispondente numero di pagina, ad esempio per creare un indice
- ▶ Dati: nome, contenuto, posizione, ecc.
- ▶ Dimensioni della pagina, CropBox, rotazione
- ▶ Stato dello standard ISO: PDF/X, PDF/A, PDF/UA, PDF/E e PDF/VT
- ▶ Riferimento geospaziale
- ▶ Elenco o estrazione degli allegati
- ▶ Nomi dei livelli, etichette, articoli
- ▶ Dettagli sulle annotazioni
- ▶ Elenco di tutti i commenti con il nome del revisore
- ▶ Dettagli sulla firma digitale: nome del campo della firma, firmato/non firmato, nome, data e motivo della firma
- ▶ Estrazione dei profili ICC dai documenti PDF/X o PDF/A

- ▶ Proprietà del Block container per il PDFlib Personalization Server
- ▶ JavaScript a livello di documento, pagina, annotazione o campo

## Output

PDFlib pCOS salva i file in diversi formati:

- ▶ Testo semplice
- ▶ Unicode UTF-8 e UTF-16
- ▶ Formato tabellare per fogli di calcolo
- ▶ Dati binari, ad es. profili ICC o allegati
- ▶ Altri formati personalizzati dall'utente

## Directory pCOS: sintassi semplice per gli oggetti PDF

Aniché dover navigare attraverso una complessa struttura ad albero, è possibile navigare all'interno della struttura del PDF, ad esempio fra i segnalibri o i campi dei moduli, semplicemente utilizzando la sintassi di pCOS Path. Grazie ai collegamenti è possibile accedere comodamente agli oggetti PDF più comuni, come le pagine, i font, i segnalibri, i campi, ecc.

## Libreria pCOS o command line?

pCOS è disponibile come libreria (componente) per numerosi ambienti di sviluppo e come command line per le operazioni di batch. Entrambe le versioni offrono le stesse funzionalità, ma sono indicate per scopi differenti.

## La libreria pCOS è indicata...

...per l'integrazione in applicazioni desktop o server. Gli esempi per utilizzare la libreria con i linguaggi di programmazione supportati sono contenuti nel pacchetto pCOS.

## La command line pCOS è indicata...

...per le operazioni di batch con documenti PDF. Non richiede nessun tipo di programmazione ma offre delle potenti opzioni che permettono di integrare l'applicazione in complessi workflow. La command line pCOS estende le funzionalità della libreria:

- ▶ Semplice recupero di elementi PDF comuni, quali segnalibri, annotazioni, metadati, campi, ecc.
- ▶ Modalità avanzata per recuperare gli oggetti complessi e personalizzare il formato dell'output
- ▶ Estrazione di elementi come allegati, profili ICC, ecc.
- ▶ Output in formato CSV o in formato definito dall'utente per importare i valori in un tabella di calcolo o in un database
- ▶ Recursione per rilevare gli oggetti PDF composti, come dictionary e array

## Ambienti di sviluppo supportati

PDFlib pCOS funziona su quasi tutte le piattaforme informatiche. Sono disponibili varianti 32-bit e 64-bit per tutte le versioni di Windows, Mac OS X, Linux e Unix.

Per garantire le migliori performance possibili e ridurre l'overhead, il pCOS core è scritto in codice C e C++ ottimizzato. Grazie ad una semplice API (Application Programming Interface) è possibile accedere alle funzionalità di pCOS da numerosi ambienti di sviluppo:

- ▶ COM per VB, ASP e molti altri linguaggi
- ▶ C e C++
- ▶ Java, servlets e Java Application Server
- ▶ .NET per C#, VB.NET, ASP.NET, ecc.
- ▶ Perl

- ▶ PHP
- ▶ Python

## Benefici di PDFlib

### Affidabile

In tutto il mondo sono migliaia i programmatori che utilizzano il nostro software. PDFlib soddisfa tutti i requisiti in termini di qualità e performance per l'utilizzo sui server. Tutti i prodotti PDFlib sono adatti per l'utilizzo continuo su server e per le operazioni di batch senza supervisione.

### Semplice e veloce

I prodotti PDFlib sono estremamente veloci e sono in grado di elaborare fino a centinaia di pagine al secondo. L'interfaccia di programmazione è intuitiva e di facile apprendimento.

### I prodotti PDFlib sono utilizzati in tutto il mondo

I nostri prodotti supportano tutte le lingue internazionali e Unicode. Vengono utilizzati da utenti in ogni angolo del mondo.

### Assistenza professionale

In caso di problemi potete contare sul nostro servizio di assistenza professionale. Per garantire che il workflow a livello aziendale non subisca mai interruzioni, offriamo la possibilità di sottoscrivere un contratto di servizio. In questo modo potrete accedere alle versioni più recenti e potrete contare su tempi di risposta celeri qualora si dovesse verificare un problema.

### Licenza

Sono a disposizione diverse tipologie di licenza: server, integrativa, aziendale e codice sorgente. È altresì possibile sottoscrivere un contratto di servizio per usufruire di un'assistenza professionale con tempi brevi di risposta e aggiornamenti gratuiti.

### L'azienda PDFlib GmbH

PDFlib GmbH è un'azienda specializzata nel settore delle tecnologie PDF. I prodotti PDFlib vengono utilizzati in tutto il mondo dal 1997. L'azienda segue attentamente gli sviluppi e i trend del mercato, come gli standard ISO per il formato PDF. PDFlib GmbH distribuisce i suoi prodotti in tutti i continenti e detiene una posizione leader nei mercati in Nordamerica, Europa e Giappone.

### Contatti

Sul nostro sito Internet sono disponibili le versioni demo complete di documentazione ed esempi. Per maggiori informazioni, contattare:



PDFlib GmbH  
 Franziska-Bilek-Weg 9, 80339 Monaco, Germania  
 Te. +49 · 89 · 452 33 84-0, Fax +49 · 89 · 452 33 84-99  
 sales@pdflib.com  
 www.pdflib.com