PDFlib pCOS
PDF Information Retrieval Tool

TELL US MORE!

PDF

What is PDFlib pCOS?

PDFlib pCOS provides a simple and elegant facility for retrieving any information from a PDF document which is not part of the page contents. For example, PDF metadata, hypertext, or page dimensions can easily be queried with pCOS.

With pCOS you can extract a variety of interesting items and create output for different purposes. By processing multiple PDF documents with a single call you can easily create summaries of document info entries, page formats, fonts, or any other property. Combined with tabular output this provides a powerful PDF administration tool.

There are many every-day pCOS applications for PDF practitioners, but you can also use PDFlib pCOS as a tool for learning or debugging PDF. Here are some typical scenarios:

► Check incoming documents for predefined criteria
► Check documents for quality assurance before publication
► Identify problem files in a large collection
► Create property summaries for document management
► Learn details of PDF data structures

Benefits of using PDFlib Software

Rock-solid Products

Tens of thousands of programmers worldwide are working with our software. PDFlib meets all quality and performance requirements for server deployment. All PDFlib products are suitable for robust 24x7 server deployment and unattended batch processing.

Speed and Simplicity

PDFlib products are incredibly fast – up to thousands of pages per second. The programming interface is straightforward and easy to learn.

PDFlib all over the World

Our products support all international languages as well as Unicode. They are used by customers in all parts of the world.

Professional Support

If there's a problem, we will try to help. We offer commercial support to meet the requirements of your business-critical applications. By adding support you will have access to the latest versions, and have guaranteed response times should any problems arise.

PDFlib pCOS Features

Supported Input

PDFlib pCOS supports all relevant flavors of PDF input:

► All PDF versions up to PDF 1.6 (Acrobat 7)
► Encrypted PDF with 40- and 128-bit encryption
► Sophisticated security model: even if you don't know the password, you can query certain pieces of information as long as this doesn't violate the document author's intentions

Information Retrieval

PDFlib pCOS offers a simple query interface, without the need for low-level parser programming. With PDFlib pCOS you can extract a variety of interesting items, such as:

► General information: linearization and tagged PDF status, encryption details and permission settings, number of pages and fonts
► Document info entries and XMP metadata
► All fonts with their name, embedding status, etc.
► Target URLs and coordinates of Web links
► Create a table of contents by extracting all bookmarks along with the corresponding page numbers
► Form field data: full field names, contents, position, etc.
► Page size, CropBox, page rotation
► Status of PDF/X compliant files
► List or extract file attachments
► Layer names
► Annotation details
► List all comments along with the reviewer's name
► Digital signature details: name of signature field(s), signed/unsigned, name of signer, date and reason of signature
► Extract ICC output intent profiles from PDF/X or PDF/A files
► List PDFlib block properties
► Find JavaScript on document, page, annotation, or field level

Output Formats

PDFlib pCOS can create output for different purposes:

► Plain text output

► Tabular output for processing with a spreadsheet or database

► Extract binary data for reuse, e.g. ICC profiles or file attachments

► Unicode text output in UTF-8 or UTF-16 formats

► User-defined output formats for custom post-processing

pCOS Paths – Simple Syntax for PDF Objects

Instead of getting bogged down by complex tree structures (e.g. bookmarks or form fields) you can easily access PDF objects by using the simple pCOS path syntax. It offers convenient shortcuts for accessing commonly used PDF objects, such as pages, fonts, bookmarks, form fields etc.

pCOS Library or Command-Line Tool?

pCOS is available as a programming library (component) for various development environments, and as a command-line tool for batch operations. Both offer similar features, but are suitable for different deployment tasks.

The pCOS programming library is used...

...for integration into your desktop or server application. Examples for using the library with all supported language bindings are included in the pCOS package.

The pCOS command-line tool is suited...

...for batch processing PDF documents. It doesn't require any programming, but offers powerful command-line options which can be used to integrate it into complex workflows. The pCOS command-line tool extends the features of the library:

► Simple retrieval of common PDF elements, such as bookmarks, annotations, metadata, form fields, etc.

► Extended mode for querying more complex objects and customizing the output format

► Emit information as comma-separated values or a user-defined format for import into a spreadsheet or database

► Recursion feature for dumping composite PDF objects, such as dictionaries and arrays

Supported Development Environments

PDFlib pCOS is everywhere – it runs on practically all computing platforms. We offer variants for all common flavors of Windows, Mac OS, Linux and Unix, as well as for IBM eServer iSeries and zSeries mainframes.

The pCOS core is written in highly optimized C code for maximum performance and small overhead. Via a simple API (Application Programming Interface) the pCOS functionality is accessible from a variety of development environments:

► COM for use with VB, ASP, Borland Delphi, etc.

► C and C++

► Java, including servlets and Java Application Server

► .NET for use with C#, VB.NET, ASP.NET, etc.

Licensing

We offer various licensing programs for server licenses, integration and site licenses, and source code licenses. Support contracts for extended technical support with short response times and free updates are also available.

About PDFlib GmbH

PDFlib GmbH is completely focused on PDF technology. PDF development is what we have been doing since 1997, and all our products are related to PDF. We have long-time experience and can afford to research even the tiniest detail related to PDF. Our products are distributed all over the world with major markets in North America, Japan and Europe.

PDFlib
®

Contact

Fully functional evaluation versions including documentation and samples are available on our Web site for all supported platforms. For more information please contact:

PDFlib GmbH

Tal 40, 80331 München, Germany

phone +49 • 89 • 29 16 46 87

fax +49 • 89 • 29 16 46 86

sales@pdflib.com

www.pdflib.com