



ActivePDF is the leading global provider of server-side PDF automation tools. We partner with developers, product managers, CIOs, and CTOs that need a single advisor to embed feature-rich digital transformation technologies directly into their enterprise applications and IIM solutions that reliably scale for millions of users.

ActivePDF Family of Products



DocGenius™

Rock solid, scalable PDF tools for Windows developers.

Compressor

PDF Compression
[.NET/COM API]

Rasterizer

PDF to IMG
[.NET/COM API]

Redactor

PDF Redaction
[.NET API]

Server

In-App PDF Generation
[.NET/COM API]

Spooler

Print PDF to Paper
[.NET/COM API]

Toolkit

Full PDF Manipulation
[.NET/COM API]

WebGrabber

HTML to PDF
[.NET/COM API]

Xtractor

PDF Data Extraction
[.NET/COM API]



DocSight™

Server-based PDF content curation automation for digital business-critical applications.

DocConverter

PDF Conversion
[Solution w/ API]

Meridian

Network PDF Printer
[Solution w/ API]

OCR

Full-Text & Zonal OCR
[Solution w/ API]



DocSpace™

A family of server applications with user interfaces for viewing, creating, and interacting with PDF.

Reader Plus

PDF Viewer/Editor
[Solution w/ API]



Xtractor

Xtractor is the developer tool that you need for searching and extracting text and images from PDF files. Quickly locate and extract PDF-based text or images for use in other applications. Simply specify the criteria (such as words, invoice data, image formats, location of interest) and the content becomes immediately available for automation, editing, indexing, and more.

[TRY IT FOR FREE](#) [CONTACT US](#)

Highlights:

Classification and Indexing

Xtractor allows you to automate key information extraction to categorize or index documents for archiving, classification, and more.

Extract and Save

Save all extracted text to a selected memory stream. Extract and save images to a specified file name by targeting all images, specified pages, or placement coordinates.

Automate PDF Processing

Easily automate high-volume search and extraction based on contextual information such as keywords, key phrases, location, and more.

Features:

- .NET/COM API
- Specify location and return text
- Extract metadata for PDF indexing and archiving
- Extract JPG, TIFF, PNG, or BMP
- Specify location and return text
- Save to a file or memory stream
- Locate and extract invisible text
- Multithreaded
- And more!

Requirements:

Operating System

- Windows Server® 2008, 2008 R2, 2012 R2, or 2016
- Microsoft® .NET Framework 3.5.1 or higher
- Microsoft Windows® 7, 8 or 10

Code Snippet:

```
// Open a PDF file or byte array
results = oXT.OpenPDF(strPath + "DOC.pdf");

// Get Text by Location
ExtractedText = oXT.GetTextByLocation(4, 100, 400, 200);

// Get location of text
AllExtractedText = oXT.FindAll("Present");

// Find one instance of text using a regular expression
ExtractedTextRegex =
oXT.FindInstanceByRegex(@"w*sque\b", 3, 1);

// Close the PDF when finished with it
results = oXT.ClosePDF();
```