

How to convert scanned PDF to text with PDF extractor SDK in C# with ByteScout Premium Suite

This code in C# shows how to convert scanned PDF to text with PDF extractor SDK with this how to tutorial

We made thousands of pre-made source code pieces for easy implementation in your own programming projects. What is ByteScout Premium Suite? It is the set that includes 12 SDK products from ByteScout including tools and components for PDF, barcodes, spreadsheets, screen video recording. It can help you to convert scanned PDF to text with PDF extractor SDK in your C# application.

This prolific sample source code in C# for ByteScout Premium Suite contains various functions and other necessary options you should do calling the API to convert scanned PDF to text with PDF extractor SDK. This C# sample code is all you need for your app. Just copy and paste the code, add references (if needs to) and you are all set! Want to see how it works with your data then code testing will allow the function to be tested and work properly.

ByteScout Premium Suite free trial version is available on our website. C# and other programming languages are supported.

FOR MORE INFORMATION AND FREE TRIAL:

[Download Free Trial SDK \(on-premise version\)](#)

[Read more about ByteScout Premium Suite](#)

[Explore API Documentation](#)

[Get Free Training for ByteScout Premium Suite](#)

[Get Free API key for Web API](#)

[visit www.Bytescout.com](http://www.Bytescout.com)

Source Code Files:

```
using System.Diagnostics;
using Bytescout.PDFExtractor;

// This example demonstrates the use of Optical Character Recognition (OCR) to extract
// from scanned PDF documents and raster images.

// To make OCR work you should add the following references to your project:
// 'Bytescout.PDFExtractor.dll', 'Bytescout.PDFExtractor.OCRExtension.dll'.

namespace ScannedPdfToText
{
    class Program
    {
        static void Main(string[] args)
        {
            // Create Bytescout.PDFExtractor.TextExtractor instance
            TextExtractor extractor = new TextExtractor();
            extractor.RegistrationName = "demo";
            extractor.RegistrationKey = "demo";

            // Load sample PDF document
            extractor.LoadDocumentFromFile("sample_ocr.pdf");

            // Enable Optical Character Recognition (OCR)
            // in .Auto mode (SDK automatically checks if needs to use OCR or not)
            extractor.OCRMode = OCRMode.Auto;

            // Set the location of OCR language data files
            extractor.OCRLanguageDataFolder = @"c:\Program Files\Bytescout PDF Extractor";

            // Set OCR language
            extractor.OCRLanguage = "eng"; // "eng" for english, "deu" for German, "fre" for French
            // Find more language files at https://github.com/bytescout/ocrdata

            // Set PDF document rendering resolution
            extractor.OCRResolution = 300;

            // You can also apply various preprocessing filters
            // to improve the recognition on low-quality scans.

            // Automatically deskew skewed scans
            //extractor.OCRImagePreprocessingFilters.AddDeskew();

            // Remove vertical or horizontal lines (sometimes helps to avoid OCR engine)
            //extractor.OCRImagePreprocessingFilters.AddVerticalLinesRemover();
            //extractor.OCRImagePreprocessingFilters.AddHorizontalLinesRemover();

            // Repair broken letters
            //extractor.OCRImagePreprocessingFilters.AddDilate();

            // Remove noise
            //extractor.OCRImagePreprocessingFilters.AddMedian();

            // Apply Gamma Correction
```

```
//extractor.OCRImagePreprocessingFilters.AddGammaCorrection();

// Add Contrast
//extractor.OCRImagePreprocessingFilters.AddContrast(20);

// (!) You can use new OCRAnalyser class to find an optimal set of image p
// filters for your specific document.
// See "OCR Analyser" example.

// Save extracted text to file
extractor.SaveTextToFile("output.txt");

// Cleanup
extractor.Dispose();

// Open result document in default associated application (for demo purpos
ProcessStartInfo processStartInfo = new ProcessStartInfo("output.txt");
processStartInfo.UseShellExecute = true;
Process.Start(processStartInfo);
    }
}
}
```

VIDEO

<https://www.youtube.com/watch?v=NEwNs2b9YN8>

ON-PREMISE OFFLINE SDK

[60 Day Free Trial](#) or [Visit ByteScout Premium Suite Home Page](#)
[Explore ByteScout Premium Suite Documentation](#)
[Explore Samples](#)
[Sign Up for ByteScout Premium Suite Online Training](#)

ON-DEMAND REST WEB API

[Get Your API Key](#)
[Explore Web API Docs](#)
[Explore Web API Samples](#)

[visit www.ByteScout.com](http://www.ByteScout.com)

[visit www.PDF.co](http://www.PDF.co)

www.bytescout.com