

scanned PDF to CSV in C# and ByteScout PDF Extractor SDK

How To: tutorial on scanned PDF to CSV in C#

Every ByteScout tool contains example C# source codes that you can find here or in the folder with installed ByteScout product. ByteScout PDF Extractor SDK was made to help with scanned PDF to CSV in C#. ByteScout PDF Extractor SDK is the SDK that helps developers to extract data from unstructured documents, pdf, images, scanned and electronic forms. Includes AI functions like automatic table detection, automatic table extraction and restructuring, text recognition and text restoration from pdf and scanned documents. Includes PDF to CSV, PDF to XML, PDF to JSON, PDF to searchable PDF functions as well as methods for low level data extraction.

Fast application programming interfaces of ByteScout PDF Extractor SDK for C# plus the instruction and the C# code below will help you quickly learn scanned PDF to CSV. This C# sample code should be copied and pasted into your application's code editor. Then just compile and run it to see how it works. Code testing will allow the function to be tested and work properly with your data.

Free trial version of ByteScout PDF Extractor SDK is available on our website. Get it to try other samples for C#.

FOR MORE INFORMATION AND FREE TRIAL:

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

[Read more about ByteScout PDF Extractor SDK](#)

[Explore API Documentation](#)

[Get Free Training for ByteScout PDF Extractor SDK](#)

[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 ScannedPdfToCSV
{
    class Program
    {
        static void Main(string[] args)
        {
            // Create Bytescout.PDFExtractor.CSVExtractor instance
            CSVExtractor extractor = new CSVExtractor();
            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.SaveCSVToFile("output.csv");

// Cleanup
extractor.Dispose();

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

VIDEO

https://www.youtube.com/watch?v=s28W3_KMrU

ON-PREMISE OFFLINE SDK

[60 Day Free Trial](#) or [Visit ByteScout PDF Extractor SDK Home Page](#)
[Explore ByteScout PDF Extractor SDK Documentation](#)
[Explore Samples](#)
[Sign Up for ByteScout PDF Extractor SDK 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