

www.bytescout.com

OCR analyser with PDF extractor SDK in C# using ByteScout Data Extraction Suite

Learn OCR analyser with PDF extractor SDK in C#

The documentation is crafted to assist you to apply the features on your side easily. ByteScout Data Extraction Suite helps with OCR analyser with PDF extractor SDK in C#. ByteScout Data Extraction Suite is the set that includes 3 SDK products for data extraction from PDF, scans, images and from spreadsheets: PDF Extractor SDK, Data Extraction SDK, Barcode Reader SDK.

If you want to quickly learn then these fast application programming interfaces of ByteScout Data Extraction Suite for C# plus the guideline and the C# code below will help you quickly learn OCR analyser with PDF extractor SDK. Follow the steps-by-step instructions from the scratch to work and copy and paste code for C# into your editor. Further improvement of the code will make it more robust.

Trial version can be downloaded from our website for free. It contains this and other source code samples for C#.

FOR MORE INFORMATION AND FREE TRIAL:

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

[Read more about ByteScout Data Extraction Suite](#)

[Explore API Documentation](#)

[Get Free Training for ByteScout Data Extraction Suite](#)

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

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

Source Code Files:

Program.cs

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

// This example demonstrates the use of OCR Analyser - a tooling class for analysis of
// in PDF or raster image formats to find best parameters for Optical Character Recognition
// provide highest recognition quality.

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

namespace OCRAnalyser
{
    class Program
    {
        static void Main(string[] args)
        {
            // Input document
            string inputDocument = @".\sample_ocr.pdf";

            // Document page index
            int pageIndex = 0;

            // Area of the document page to perform the analysis (optional).
            // RectangleF.Empty means the full page.
            RectangleF rectangle = RectangleF.Empty; // new RectangleF(100, 50, 350, 250);

            // Location of language data files
            string ocrLanguageDataFolder = @"c:\Program Files\Bytescout PDF Extractor\Languages\en-US\";
            string ocrLanguage = "eng"; // "eng" for english, "deu" for German, "fra" for French
            // Find more language files at https://github.com/bytescout/ocrdata

            // Create OCRAnalyzer instance and activate it with your registration info
            using (OCRAnalyzer ocrAnalyzer = new OCRAnalyzer("demo", "demo"))
            {
                // Display analysis progress
                ocrAnalyzer.ProgressChanged += (object sender, string message, double progress)
                {
                    Console.WriteLine(message);
                };

                // Load document to OCRAnalyzer
                ocrAnalyzer.LoadDocumentFromFile(inputDocument);

                // Setup OCRAnalyzer
                ocrAnalyzer.OCRLanguage = ocrLanguage;
                ocrAnalyzer.OCRLanguageDataFolder = ocrLanguageDataFolder;

                // Set page area for analysis (optional)
                ocrAnalyzer.SetExtractionArea(rectangle);

                // Perform analysis and get results
                OCRAnalysisResults analysisResults = ocrAnalyzer.AnalyzeByOCRConfidence();
            }
        }
    }
}
```

```
// Now extract the text using detected OCR parameters

string outputDocument = @".\result.txt";

// Create TextExtractor instance
using (TextExtractor textExtractor = new TextExtractor("demo", "demo"))
{
    // Load document to TextExtractor
    textExtractor.LoadDocumentFromFile(inputDocument);

    // Setup TextExtractor
    textExtractor.OCRMode = OCRMode.Auto;
    textExtractor.OCRLanguageDataFolder = ocrLanguageDataFolder;
    textExtractor.OCRLanguage = ocrLanguage;

    // Apply analysis results to TextExtractor instance
    ocrAnalyzer.ApplyResults(analysisResults, textExtractor);

    // Set extraction area (optional)
    textExtractor.SetExtractionArea(rectangle);

    // Save extracted text to file
    textExtractor.SaveTextToFile(outputDocument);

    // Open result document in default associated application (for demo)
    ProcessStartInfo processStartInfo = new ProcessStartInfo(outputDocument);
    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 Data Extraction Suite Home Page](#)
[Explore ByteScout Data Extraction Suite Documentation](#)
[Explore Samples](#)

Sign Up for ByteScout Data Extraction Suite Online Training

ON-DEMAND REST WEB API

[Get Your API Key](#)

[Explore Web API Docs](#)

[Explore Web API Samples](#)

[visit www.ByteScout.com](#)

[visit www.PDF.co](#)

[www.bytescout.com](#)