

ocr with best dataset with pdf extractor sdk in VB.NET with ByteScout Data Extraction Suite

Simple tutorial on how to do ocr with best dataset with pdf extractor sdk in VB.NET

ByteScout tutorials describe the stuff for programmers who use VB.NET. ByteScout Data Extraction Suite helps with ocr with best dataset with pdf extractor sdk in VB.NET. ByteScout Data Extraction Suite is the bundle that includes three SDK tools for data extraction from PDF, scans, images and from spreadsheets: PDF Extractor SDK, Data Extraction SDK, Barcode Reader SDK.

The below SDK samples describe how to quickly make your application do ocr with best dataset with pdf extractor sdk in VB.NET with the help of ByteScout Data Extraction Suite. If you want to know how it works, then this VB.NET sample code should be copied and pasted into your application's code editor. Then just compile and run it. This basic programming language sample code for VB.NET will do the whole work for you in implementing ocr with best dataset with pdf extractor sdk in your app.

If you want to try other samples for VB.NET then free trial version of ByteScout Data Extraction Suite is available on our website.

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:

```
Imports Bytescout.PDFExtractor
```

```
' This example demonstrates the use of Optical Character Recognition (OCR) to extract text
' 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".
```

```
Class Program
```

```
    Friend Shared Sub Main(args As String())
```

```
        ' Create Bytescout.PDFExtractor.TextExtractor instance
```

```
        Dim extractor As 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 SDK\Languages"
```

```
        ' Set OCR language
```

```
        extractor.OCRLanguage = "eng" ' "eng" for english, "deu" for German, "fra" for French
```

```
        ' Find more language files at https://github.com/bytescout/ocrdata/tree/master
```

```
        ' 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's problems)
```

```
        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 OCRAnalyzer class to find an optimal set of image preproc
' filters for your specific document.
' See "OCR Analyser" example.

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

' Cleanup
extractor.Dispose()

' Open output file in default associated application
System.Diagnostics.Process.Start("output.txt")

End Sub

End Class
```

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

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

