

How to convert scanned pdf to json with pdf extractor sdk in VB.NET using ByteScout PDF Suite

Learn to convert scanned pdf to json with pdf extractor sdk in VB.NET

An easy to understand sample source code to learn how to convert scanned pdf to json with pdf extractor sdk in VB.NET ByteScout PDF Suite is the bundle that provides six different SDK libraries to work with PDF from generating rich PDF reports to extracting data from PDF documents and converting them to HTML. This bundle includes PDF (Generator) SDK, PDF Renderer SDK, PDF Extractor SDK, PDF to HTML SDK, PDF Viewer SDK and PDF Generator SDK for Javascript and you can use it to convert scanned pdf to json with pdf extractor sdk with VB.NET.

These VB.NET code samples for VB.NET guide developers to speed up coding of the application when using ByteScout PDF Suite. IF you want to implement the functionality, just copy and paste this code for VB.NET below into your code editor with your app, compile and run your application. Use of ByteScout PDF Suite in VB.NET is also described in the documentation included along with the product.

Our website gives trial version of ByteScout PDF Suite for free. It also includes documentation and source code samples.

FOR MORE INFORMATION AND FREE TRIAL:

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

[Read more about ByteScout PDF Suite](#)

[Explore API Documentation](#)

[Get Free Training for ByteScout PDF 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.JSONExtractor instance
```

```
        Dim extractor As New JSONExtractor()
```

```
        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
```

```
        ' 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.OCRIImagePreprocessingFilters.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.SaveJSONToFile("output.json")

' Cleanup
extractor.Dispose()

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

End Sub

End Class
```

VIDEO

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

ON-PREMISE OFFLINE SDK

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

