

How to convert scanned pdf to json with pdf extractor sdk in VB.NET with ByteScout Data Extraction Suite

Step-by-step tutorial on how to convert scanned pdf to json with pdf extractor sdk in VB.NET

The sample source codes on this page shows how to convert scanned pdf to json 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 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 Data Extraction Suite. Just copy and paste the code into your VB.NET application's code and follow the instructions. If you want to use these VB.NET sample examples in one or many applications then they can be used easily.

You can download free trial version of ByteScout Data Extraction Suite from our website with this and other source code samples for VB.NET.

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.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 prepro  
' 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 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)

