

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

If you want to learn more then this tutorial will show how to convert scanned pdf to xml with pdf extractor sdk in VB.NET

Every ByteScout tool includes simple example VB.NET source codes that you can get here or in the folder with installed ByteScout product. ByteScout Data Extraction Suite can convert scanned pdf to xml with pdf extractor sdk. It can be applied from VB.NET. 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.

This prolific sample source code in VB.NET for ByteScout Data Extraction Suite contains various functions and other necessary options you should do calling the API to convert scanned pdf to xml with pdf extractor sdk. Follow the instructions from scratch to work and copy the VB.NET code. If you want to use these VB.NET sample examples in one or many applications then they can be used easily.

ByteScout Data Extraction Suite free trial version is available on our website. VB.NET and other programming languages are supported.

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.XMLExtractor instance
```

```
        Dim extractor As New XMLExtractor()  
        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"
```

```
        ' 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.OCRImagePreprocessingFilters.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.SaveXMLToFile("output.xml")
```

```
' Cleanup  
extractor.Dispose()
```

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

```
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)

