

PDF to XML with images in VB.NET using ByteScout PDF Extractor SDK

Tutorial: how to do PDF to XML with images in VB.NET

Source code documentation samples provide quick and easy way to add a required functionality into your application. ByteScout PDF Extractor SDK was made to help with PDF to XML with images in VB.NET. ByteScout PDF Extractor SDK is the Software Development Kit (SDK) that is designed to help developers with data extraction from unstructured documents like pdf, tiff, scans, images, scanned and electronic forms. The library is powered by OCR, computer vision and AI to provide unique functionality like table detection, automatic table structure extraction, data restoration, data restructuring and reconstruction. Supports PDF, TIFF, PNG, JPG images as input and can output CSV, XML, JSON formatted data. Includes full set of utilities like pdf splitter, pdf merger, searchable pdf maker.

The SDK samples like this one below explain how to quickly make your application do PDF to XML with images in VB.NET with the help of ByteScout PDF Extractor SDK. In order to implement this functionality, you should copy and paste code below into your app using code editor. Then compile and run your application. VB.NET application implementation typically includes multiple stages of the software development so even if the functionality works please test it with your data and the production environment.

ByteScout PDF Extractor SDK free trial version is available for download from our website. Free trial also includes programming tutorials along with source code samples.

FOR MORE INFORMATION AND FREE TRIAL:

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

[Read more about ByteScout PDF Extractor SDK](#)

[Explore API Documentation](#)

[Get Free Training for ByteScout PDF Extractor SDK](#)

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

[visit www.Bytescout.com](http://www.Bytescout.com)

Source Code Files:

Module1.vb

```
Imports Bytescout.PDFExtractor

Namespace PDF2XML

    Class Program

        Shared Sub Main(ByVal 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("sample1.pdf")

            ' Uncomment this line to get rid of empty nodes in XML
            'extractor.PreserveFormattingOnTextExtraction = False

            ' Set output image format
            extractor.ImageFormat = OutputImageFormat.PNG

            ' Save images to external files
            extractor.SaveImages = ImageHandling.OuterFile
            extractor.ImageFolder = "images" ' Folder for external images
            extractor.SaveXMLToFile("result_with_external_images.xml")

            ' Embed images into XML as Base64 encoded string
            extractor.SaveImages = ImageHandling.Embed
            extractor.SaveXMLToFile("result_with_embedded_images.xml")

            ' Cleanup
            extractor.Dispose()

        End Sub

    End Class

End Namespace
```

VIDEO

https://www.youtube.com/watch?v=s28W3_KMraU

ON-PREMISE OFFLINE SDK

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

www.bytescout.com