How to find table in PDF and extract as XML in VB.NET with ByteScout PDF Extractor SDK

How to find table in PDF and extract as XML in VB.NET

Every ByteScout tool contains example VB.NET source codes that you can find here or in the folder with installed ByteScout product. Want to find table in PDF and extract as XML in your VB.NET app? ByteScout PDF Extractor SDK is designed for it. 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.

You will save a lot of time on writing and testing code as you may just take the VB.NET code from ByteScout PDF Extractor SDK for find table in PDF and extract as XML below and use it in your application. This VB.NET sample code is all you need for your app. Just copy and paste the code, add references (if needs to) and you are all set! You can use these VB.NET sample examples in one or many applications.

ByteScout free trial version is available for download from our website. It includes all these 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

Source Code Files:

Program.vb

```vb
Imports Bytescout.PDFExtractor

Class Program
    Friend Shared Sub Main(args As String())

        ' Create Bytescout.PDFExtractor.XMLExtractor instance
        Dim xmlExtractor As New XMLExtractor()
        xmlExtractor.RegistrationName = "demo"
        xmlExtractor.RegistrationKey = "demo"

        ' Create Bytescout.PDFExtractor.TableDetector instance
        Dim tableDetector As New TableDetector()
        tableDetector.RegistrationName = "demo"
        tableDetector.RegistrationKey = "demo"

        ' We should define what kind of tables we should detect.
        ' So we set min required number of columns to 3 ...
        tableDetector.DetectionMinNumberOfColumns = 3
        ' ... and we set min required number of rows to 3
        tableDetector.DetectionMinNumberOfRows = 3

            ' Load sample PDF document
        xmlExtractor.LoadDocumentFromFile(".\sample3.pdf")
        tableDetector.LoadDocumentFromFile(".\sample3.pdf")

            ' Get page count
        Dim pageCount As Integer = tableDetector.GetPageCount()

            For i As Integer = 0 To pageCount - 1
        Dim t As Integer = 1
        ' Find first table and continue if found
        If (tableDetector.FindTable(i)) Then
            Do
                ' Set extraction area for XML extractor to rectangle received from
                xmlExtractor.SetExtractionArea(tableDetector.FoundTableLocation)
                ' Export the table to XML file
                xmlExtractor.SavePageXMLToFile(i, "page-" + i.ToString() + "-table-
                t = t + 1
            Loop While tableDetector.FindNextTable()
        End If
        Next

        ' Cleanup
            xmlExtractor.Dispose()
            tableDetector.Dispose()

        ' Open first output file in default associated application (for demo purposes)
        System.Diagnostics.Process.Start("page-0-table-1.xml")

    End Sub
```

```
End Class
```

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

visit www.PDF.co