

How to find PDF table and extract as XML in VBScript and ByteScout PDF Extractor SDK

Tutorial on how to find PDF table and extract as XML in VBScript

Learn how to find PDF table and extract as XML in VBScript with this source code sample. Want to find PDF table and extract as XML in your VBScript app? ByteScout PDF Extractor SDK is designed for it. ByteScout PDF Extractor SDK is the SDK is designed to help developers with pdf tables and pdf 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 and other utilities.

VBScript code samples for VBScript developers help to speed up coding of your application when using ByteScout PDF Extractor SDK. In your VBScript project or application you may simply copy & paste the code and then run your app! You can use these VBScript sample examples in one or many applications.

Trial version of ByteScout PDF Extractor SDK can be downloaded for free from our website. It also includes source code samples for VBScript and other programming languages.

VBScript - FindTableAndExtractAsXML.vbs

```
' Create Bytescout.PDFExtractor.TextExtractor object
Set tableDetector= CreateObject("Bytescout.PDFExtractor.TableDetector")
tableDetector.RegistrationName = "demo"
tableDetector.RegistrationKey = "demo"

' Create Bytescout.PDFExtractor.xmlExtractor object
Set xmlExtractor = CreateObject("Bytescout.PDFExtractor.XMLExtractor")
xmlExtractor.RegistrationName = "demo"
xmlExtractor.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
tableDetector.LoadDocumentFromFile("../..\sample3.pdf")
xmlExtractor.LoadDocumentFromFile "../..\sample3.pdf"

' Get page count
pageCount = tableDetector.GetPageCount()
```

```

' Iterate through pages
For i = 0 to pageCount - 1

    t = 0
    ' Find first table and continue if found
    If (tableDetector.FindTable(i)) Then

        Do
            ' Set extraction area for CSV extractor to rectangle received
            from the table detector
            xmlExtractor.SetExtractionArea _
                tableDetector.GetFoundTableRectangle_Left(), _
                tableDetector.GetFoundTableRectangle_Top(), _
                tableDetector.GetFoundTableRectangle_Width(), _
                tableDetector.GetFoundTableRectangle_Height()
            ' Export the table to CSV file
            xmlExtractor.SavePageXMLToFile i, "page-" & CStr(i) & "-
            table-" & CStr(t) & ".xml"
            t = t + 1
            Loop While tableDetector.FindNextTable()

        End If

    Next

Set xmlExtractor = Nothing
Set tableDetector = Nothing

```

FOR MORE INFORMATION AND FREE TRIAL:

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

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

[Explore documentation](#)

[Visit www.ByteScout.com](#)

or

[Get Your Free API Key for www.PDF.co Web API](#)