How to find PDF table and extract as XML in C# with ByteScout PDF Extractor SDK

How to find PDF table and extract as XML in C#

Every ByteScout tool contains example C# source codes that you can find here or in the folder with installed ByteScout product. ByteScout PDF Extractor SDK can find PDF table and extract as XML. It can be used from C#. ByteScout PDF Extractor SDK is the SDK that helps developers to extract data from unstructured documents, pdf, images, scanned and electronic forms. Includes AI functions like automatic table detection, automatic table extraction and restructuring, text recognition and text restoration from pdf and scanned documents. Includes PDF to CSV, PDF to XML, PDF to JSON, PDF to searchable PDF functions as well as methods for low level data extraction.

Fast application programming interfaces of ByteScout PDF Extractor SDK for C# plus the instruction and the code below will help you quickly learn how to find PDF table and extract as XML. This C# 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! Use of ByteScout PDF Extractor SDK in C# is also explained in the documentation included along with the product.

Free trial version of ByteScout PDF Extractor SDK is available for download from our website. Get it to try other source code samples for C#.

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.cs

```csharp
using System.Diagnostics;
using Bytescout.PDFExtractor;

namespace FindTableAndExtractAsXml
{
    class Program
    {
        static void Main(string[] args)
        {
            // Create Bytescout.PDFExtractor.XMLExtractor instance
            XMLExtractor xmlExtractor = new XMLExtractor();
            xmlExtractor.RegistrationName = "demo";
            xmlExtractor.RegistrationKey = "demo";

            // Create Bytescout.PDFExtractor.TableDetector instance
            TableDetector tableDetector = new TableDetector();
            tableDetector.RegistrationKey = "demo";
            tableDetector.RegistrationName = "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
            int pageCount = tableDetector.GetPageCount();

            for (int i = 0; i < pageCount; i++)
            {
                int t = 1;
                // Find first table and continue if found
                if (tableDetector.FindTable(i))
                {
                    do
                    {
                        // Set extraction area for XML extractor to rectangle received
                        xmlExtractor.SetExtractionArea(tableDetector.FoundTableLocation
                        // Export the table to XML file
                        xmlExtractor.SavePageXMLToFile(i, "page-" + i + "-table-" + t +
                        t++;
                    }
                    while (tableDetector.FindNextTable()); // search next table
                }
            }

            // Cleanup
            xmlExtractor.Dispose();
```

```
            tableDetector.Dispose();

            // Open first output file in default associated application (for demo purp
            ProcessStartInfo processStartInfo = new ProcessStartInfo("page-0-table-1.xr
            processStartInfo.UseShellExecute = true;
            Process.Start(processStartInfo);
        }
    }
}
```

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