

www.bytescout.com

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

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.

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:

FindTableAndExtractAsJson.VS2005.vbproj

```
<?xml version="1.0" encoding="utf-8"?>
<Project DefaultTargets="Build" xmlns="http://schemas.microsoft.com/developer/msbuild/
  <PropertyGroup>
    <Configuration Condition=" '$(Configuration)' == '' ">Debug</Configuration>
    <Platform Condition=" '$(Platform)' == '' ">AnyCPU</Platform>
    <ProductVersion>8.0.50727</ProductVersion>
    <SchemaVersion>2.0</SchemaVersion>
    <ProjectGuid>{34509168-5D95-4323-8808-2A10FDE4E9A9}</ProjectGuid>
    <OutputType>Exe</OutputType>
    <AppDesignerFolder>Properties</AppDesignerFolder>
    <RootNamespace>FindTableAndExtractAsJson</RootNamespace>
    <AssemblyName>FindTableAndExtractAsJson</AssemblyName>
  </PropertyGroup>
  <PropertyGroup Condition=" '$(Configuration)|$(Platform)' == 'Debug|AnyCPU' ">
    <DebugSymbols>>true</DebugSymbols>
    <DebugType>full</DebugType>
    <Optimize>>false</Optimize>
    <OutputPath>bin\Debug\</OutputPath>
    <DefineConstants>DEBUG,TRACE</DefineConstants>
    <ErrorReport>prompt</ErrorReport>
    <WarningLevel>4</WarningLevel>
  </PropertyGroup>
  <PropertyGroup Condition=" '$(Configuration)|$(Platform)' == 'Release|AnyCPU' ">
    <DebugType>pdbonly</DebugType>
    <Optimize>>true</Optimize>
    <OutputPath>bin\Release\</OutputPath>
    <DefineConstants>TRACE</DefineConstants>
    <ErrorReport>prompt</ErrorReport>
    <WarningLevel>4</WarningLevel>
  </PropertyGroup>
  <Import Project="$(MSBuildBinPath)\Microsoft.VisualBasic.Targets" />
  <ItemGroup>
    <Import Include="Microsoft.VisualBasic" />
    <Import Include="System" />
    <Reference Include="Bytescout.PDFExtractor, Version=1.0.0.12, Culture=neutral, pro
      <SpecificVersion>False</SpecificVersion>
    </Reference>
    <Reference Include="System" />
    <Reference Include="System.Data" />
    <Reference Include="System.Drawing" />
    <Reference Include="System.Xml" />
    <Compile Include="Program.vb" />
    <Compile Include="Properties\AssemblyInfo.vb" />
  </ItemGroup>
  <ItemGroup>
    <Content Include="sample3.pdf">
      <CopyToOutputDirectory>Always</CopyToOutputDirectory>
    </Content>
  </ItemGroup>
</Project>
```

FindTableAndExtractAsJson.VS2008.vbproj

```
<?xml version="1.0" encoding="utf-8"?>
<Project DefaultTargets="Build" xmlns="http://schemas.microsoft.com/developer/msbuild/
  <PropertyGroup>
    <Configuration Condition=" '$(Configuration)' == '' ">Debug</Configuration>
    <Platform Condition=" '$(Platform)' == '' ">AnyCPU</Platform>
    <ProductVersion>9.0.21022</ProductVersion>
    <SchemaVersion>2.0</SchemaVersion>
    <ProjectGuid>{34509168-5D95-4323-8808-2A10FDE4E9A9}</ProjectGuid>
    <OutputType>Exe</OutputType>
    <AppDesignerFolder>Properties</AppDesignerFolder>
    <RootNamespace>FindTableAndExtractAsJson</RootNamespace>
    <AssemblyName>FindTableAndExtractAsJson</AssemblyName>
    <OldToolsVersion>2.0</OldToolsVersion>
    <TargetFrameworkVersion>v3.5</TargetFrameworkVersion>
  </PropertyGroup>
  <PropertyGroup Condition=" '$(Configuration)|$(Platform)' == 'Debug|AnyCPU' ">
    <DebugSymbols>>true</DebugSymbols>
    <DebugType>full</DebugType>
    <Optimize>>false</Optimize>
    <OutputPath>bin\Debug\</OutputPath>
    <DefineConstants>DEBUG,TRACE</DefineConstants>
    <ErrorReport>prompt</ErrorReport>
    <WarningLevel>4</WarningLevel>
  </PropertyGroup>
  <PropertyGroup Condition=" '$(Configuration)|$(Platform)' == 'Release|AnyCPU' ">
    <DebugType>pdbonly</DebugType>
    <Optimize>>true</Optimize>
    <OutputPath>bin\Release\</OutputPath>
    <DefineConstants>TRACE</DefineConstants>
    <ErrorReport>prompt</ErrorReport>
    <WarningLevel>4</WarningLevel>
  </PropertyGroup>
  <Import Project="$(MSBuildToolsPath)\Microsoft.VisualBasic.Targets" />
  <ItemGroup>
    <Import Include="Microsoft.VisualBasic" />
    <Import Include="System" />
    <Reference Include="Bytescout.PDFExtractor, Version=1.0.0.12, Culture=neutral, pro
      <SpecificVersion>False</SpecificVersion>
    </Reference>
    <Reference Include="System" />
    <Reference Include="System.Data" />
    <Reference Include="System.Drawing" />
    <Reference Include="System.Xml" />
    <Compile Include="Program.vb" />
    <Compile Include="Properties\AssemblyInfo.vb" />
  </ItemGroup>
  <ItemGroup>
    <Content Include="sample3.pdf">
      <CopyToOutputDirectory>Always</CopyToOutputDirectory>
    </Content>
  </ItemGroup>
</Project>
```

FindTableAndExtractAsJson.VS2010.vbproj

```
<?xml version="1.0" encoding="utf-8"?>
<Project DefaultTargets="Build" xmlns="http://schemas.microsoft.com/developer/msbuild/
  <PropertyGroup>
    <Configuration Condition=" '$(Configuration)' == '' ">Debug</Configuration>
    <Platform Condition=" '$(Platform)' == '' ">AnyCPU</Platform>
    <SchemaVersion>2.0</SchemaVersion>
    <ProjectGuid>{34509168-5D95-4323-8808-2A10FDE4E9A9}</ProjectGuid>
    <OutputType>Exe</OutputType>
    <AppDesignerFolder>Properties</AppDesignerFolder>
    <RootNamespace>FindTableAndExtractAsJson</RootNamespace>
    <AssemblyName>FindTableAndExtractAsJson</AssemblyName>
    <OldToolsVersion>3.5</OldToolsVersion>
    <TargetFrameworkVersion>v4.0</TargetFrameworkVersion>
  </PropertyGroup>
  <PropertyGroup Condition=" '$(Configuration)|$(Platform)' == 'Debug|AnyCPU' ">
    <DebugSymbols>>true</DebugSymbols>
    <DebugType>full</DebugType>
    <Optimize>>false</Optimize>
    <OutputPath>bin\Debug\</OutputPath>
    <DefineConstants>DEBUG,TRACE</DefineConstants>
    <ErrorReport>prompt</ErrorReport>
    <WarningLevel>4</WarningLevel>
  </PropertyGroup>
  <PropertyGroup Condition=" '$(Configuration)|$(Platform)' == 'Release|AnyCPU' ">
    <DebugType>pdbonly</DebugType>
    <Optimize>>true</Optimize>
    <OutputPath>bin\Release\</OutputPath>
    <DefineConstants>TRACE</DefineConstants>
    <ErrorReport>prompt</ErrorReport>
    <WarningLevel>4</WarningLevel>
  </PropertyGroup>
  <PropertyGroup>
    <StartupObject>Sub Main</StartupObject>
  </PropertyGroup>
  <Import Project="$(MSBuildToolsPath)\Microsoft.VisualBasic.Targets" />
  <ItemGroup>
    <Import Include="Microsoft.VisualBasic" />
    <Import Include="System" />
    <Reference Include="Bytescout.PDFExtractor, Version=1.0.0.12, Culture=neutral, pro
      <SpecificVersion>False</SpecificVersion>
    </Reference>
    <Reference Include="System" />
    <Reference Include="System.Data" />
    <Reference Include="System.Drawing" />
    <Reference Include="System.Xml" />
    <Compile Include="Program.vb" />
    <Compile Include="Properties\AssemblyInfo.vb" />
  </ItemGroup>
  <ItemGroup>
    <Content Include="sample3.pdf">
      <CopyToOutputDirectory>Always</CopyToOutputDirectory>
    </Content>
  </ItemGroup>
</Project>
```

```
</ItemGroup>  
</Project>
```

Program.vb

```
Imports Bytescout.PDFExtractor  
  
Class Program  
    Friend Shared Sub Main(args As String())  
  
        ' Create Bytescout.PDFExtractor.CSVExtractor instance  
        Dim jsonExtractor As New JSONExtractor()  
        jsonExtractor.RegistrationName = "demo"  
        jsonExtractor.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  
  
        ' Set table detection mode to "bordered tables" - best for tables with closed  
        tableDetector.ColumnDetectionMode = ColumnDetectionMode.BorderedTables  
  
        ' Load sample PDF document  
        jsonExtractor.LoadDocumentFromFile(".\sample3.pdf")  
        tableDetector.LoadDocumentFromFile(".\sample3.pdf")  
  
        ' Get page count  
        Dim pageCount As Integer = tableDetector.GetPageCount()  
  
        ' Iterate through pages  
        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 JSON extractor to rectangle received from  
                    jsonExtractor.SetExtractionArea(tableDetector.FoundTableLocation)  
                    ' Export the table to JSON file  
                    jsonExtractor.SaveJSONToFile(i, "page-" + i.ToString() + "-table-"  
                    t = t + 1  
                Loop While tableDetector.FindNextTable()  
            End If  
        Next  
    End Sub  
End Class
```

```
' Cleanup
jsonExtractor.Dispose()
tableDetector.Dispose()

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

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](http://www.ByteScout.com)

[visit www.PDF.co](http://www.PDF.co)

