

How to find invoice total amount in PDF with regex in VB.NET and ByteScout PDF Extractor SDK

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.

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:

FindInvoiceTotalNumberWithRegexp.vbproj

```
<?xml version="1.0" encoding="utf-8"?>
<Project ToolsVersion="15.0" xmlns="http://schemas.microsoft.com/developer/msbuild/2003">
  <Import Project="$(MSBuildExtensionsPath)\$(MSBuildToolsVersion)\Microsoft.Common.props" Condition="Exists('$(MSBuildExtensionsPath)\$(MSBuildToolsVersion)\Microsoft.Common.props')" />
  <PropertyGroup>
    <Configuration Condition="'$(Configuration)' == ''">Debug</Configuration>
    <Platform Condition="'$(Platform)' == ''">AnyCPU</Platform>
    <ProjectGuid>{80667702-F68F-42E8-AF48-A3F9D8C879CF}</ProjectGuid>
    <OutputType>Exe</OutputType>
    <StartupObject>FindInvoiceTotalNumberWithRegexp.Program</StartupObject>
    <RootNamespace>FindInvoiceTotalNumberWithRegexp</RootNamespace>
    <AssemblyName>FindInvoiceTotalNumberWithRegexp</AssemblyName>
  </PropertyGroup>
</Project>
```

```

<FileAlignment>512</FileAlignment>
<MyType>Console</MyType>
<TargetFrameworkVersion>v2.0</TargetFrameworkVersion>
</PropertyGroup>
<PropertyGroup Condition=" '$(Configuration)|$(Platform)' == 'Debug|AnyCPU' ">
  <PlatformTarget>AnyCPU</PlatformTarget>
  <DebugSymbols>>true</DebugSymbols>
  <DebugType>full</DebugType>
  <DefineDebug>>true</DefineDebug>
  <DefineTrace>>true</DefineTrace>
  <OutputPath>bin\Debug\</OutputPath>
  <NoWarn>42016,41999,42017,42018,42019,42032,42036,42020,42021,42022</NoWarn>
</PropertyGroup>
<PropertyGroup Condition=" '$(Configuration)|$(Platform)' == 'Release|AnyCPU' ">
  <PlatformTarget>AnyCPU</PlatformTarget>
  <DebugType>pdbonly</DebugType>
  <DefineDebug>>false</DefineDebug>
  <DefineTrace>>true</DefineTrace>
  <Optimize>>true</Optimize>
  <OutputPath>bin\Release\</OutputPath>
  <NoWarn>42016,41999,42017,42018,42019,42032,42036,42020,42021,42022</NoWarn>
</PropertyGroup>
<PropertyGroup>
  <OptionExplicit>On</OptionExplicit>
</PropertyGroup>
<PropertyGroup>
  <OptionCompare>Binary</OptionCompare>
</PropertyGroup>
<PropertyGroup>
  <OptionStrict>Off</OptionStrict>
</PropertyGroup>
<PropertyGroup>
  <OptionInfer>On</OptionInfer>
</PropertyGroup>
<ItemGroup>
  <Reference Include="Bytescout.PDFExtractor, Version=9.1.0.3170, Culture=neutral, PublicKeyToken=f7dd1bd9d40" />
    <SpecificVersion>False</SpecificVersion>
    <HintPath>c:\Program Files\Bytescout PDF Extractor SDK\net2.00\Bytescout.PDFExtractor.dll</HintPath>
  </Reference>
  <Reference Include="System" />
  <Reference Include="System.Data" />
  <Reference Include="System.Deployment" />
  <Reference Include="System.Xml" />
</ItemGroup>
<ItemGroup>
  <Import Include="Microsoft.VisualBasic" />
  <Import Include="System" />
  <Import Include="System.Collections" />
  <Import Include="System.Collections.Generic" />
  <Import Include="System.Data" />
  <Import Include="System.Diagnostics" />
</ItemGroup>
<ItemGroup>
  <Compile Include="Program.vb" />
</ItemGroup>
<ItemGroup>
  <None Include="SampleInvoice.pdf">
    <CopyToOutputDirectory>Always</CopyToOutputDirectory>
  </None>
</ItemGroup>
<Import Project="$(MSBuildToolsPath)\Microsoft.VisualBasic.targets" />
</Project>

```

Program.vb

```
Imports Bytescout.PDFExtractor
```

' Note: if you are looking for a more higher level API to extract data from invoices, reports, statements
' then please [check Document Parser SDK](https://bytescout.com/products/developer/documentparser) and Web API [at https://pdf.co](https://pdf.co) for secure and scalable web api

Module Program

Sub Main()

Try

' Create Bytescout.PDFExtractor.TextExtractor instance
Using extractor As TextExtractor = New TextExtractor()
extractor.RegistrationName = "demo"
extractor.RegistrationKey = "demo"

' Load sample PDF document
extractor.LoadDocumentFromFile("SampleInvoice.pdf")

' Enable the regular expression
extractor.RegexSearch = True

Dim pageCount As Integer = extractor.GetPageCount()

' Search through pages

For i As Integer = 0 To pageCount - 1

' Search total number

Dim regexPattern As String = "(TOTAL:)[\$]\d+"

' See the complete regular expressions reference at [https://msdn.microsoft.com/en-us/library/az24scfc\(v=vs.110\).aspx](https://msdn.microsoft.com/en-us/library/az24scfc(v=vs.110).aspx)

' Search each page for the pattern

If extractor.Find(i, regexPattern, False) Then

Do

' Iterate through each element in the found text

For Each element As ISearchResultElement In extractor.FoundText.Elements

Console.WriteLine("Found Total Amount Number: " & element.Text.Replace("TOTAL:", "").Trim())

Next

Loop While extractor.FindNext()

End If

Next

End Using

Catch ex As Exception

Console.WriteLine("Error: " & ex.Message)

End Try

Console.WriteLine()

Console.WriteLine("Press enter key to continue...")

Console.ReadLine()

End Sub

End Module

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