How to find text in PDF with regex in VB.NET with ByteScout PDF Extractor SDK

This code in VB.NET shows how to find text in PDF with regex with this how to tutorial

The sample shows steps and algorithm of how to find text in PDF with regex and how to make it work in your VB.NET application. 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. It can find text in PDF with regex in VB.NET.

The SDK samples like this one below explain how to quickly make your application do find text in PDF with regex in VB.NET with the help of ByteScout PDF Extractor SDK. Just copy and paste the code into your VB.NET application's code and follow the instruction. Implementing VB.NET application typically includes multiple stages of the software development so even if the functionality works please test it with your data and the production environment.

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

VB.NET - Program.vb

```
Imports System.Drawing
Imports Bytescout.PDFExtractor

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

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

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

        extractor.RegexSearch = True ' Enable the regular expressions

        Dim pageCount As Integer = extractor.GetPageCount()

        ' Search through pages
        For i As Integer = 0 To pageCount - 1
```

```
            ' Search dates in format 12/31/1999
            Dim regexPattern As String = "[0-9]{2}/[0-9]{2}/[0-9]{4}"
            ' See the complete regular expressions reference at
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
                    Console.WriteLine("")
                    Console.WriteLine(("Found on page " & i & " at location ") +
extractor.FoundText.Bounds.ToString())
                    Console.WriteLine("")

                    ' Iterate through each element in the found text
                    For Each element As ISearchResultElement In
extractor.FoundText.Elements
                        Console.WriteLine("   Text: " + element.Text)
                        Console.WriteLine("   Font is bold: " +
element.FontIsBold.ToString())
                        Console.WriteLine("   Font is italic:" +
element.FontIsItalic.ToString())
                        Console.WriteLine("   Font name: " + element.FontName)
                        Console.WriteLine("   Font size:" +
element.FontSize.ToString())
                        Console.WriteLine("   Font color:" +
element.FontColor.ToString())
                        Console.WriteLine()
                    Next

                Loop While extractor.FindNext()

            End If
        Next

        ' Cleanup
                extractor.Dispose()

        Console.WriteLine()
        Console.WriteLine("Press any key to continue...")
        Console.ReadLine()
    End Sub

End Class
```

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