

How to index PDF files in C# with ByteScout PDF Extractor SDK

This tutorial will show how to index PDF files in C#

With this source code sample you may quickly learn how to index PDF files in C#. What is ByteScout PDF Extractor SDK? It 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. It can help you to index PDF files in your C# application.

The SDK samples like this one below explain how to quickly make your application do index PDF files in C# with the help of ByteScout PDF Extractor SDK. Follow the instructions from the scratch to work and copy the C# code. Implementing C# 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.

You can download free trial version of ByteScout PDF Extractor SDK from our website to see and try many others source code samples for C#.

C# - Program.cs

```
using System;
using System.IO;
using Bytescout.PDFExtractor;

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

            TextExtractor textExtractor = new TextExtractor();
            textExtractor.RegistrationName = "demo";
            textExtractor.RegistrationKey = "demo";

            // List all PDF files in directory
            foreach (string file in Directory.GetFiles(@"..\..\..\.",
            "*.pdf"))
            {
```

```

        infoExtractor.LoadDocumentFromFile(file);

        Console.WriteLine("File Name: " +
Path.GetFileName(file));
        Console.WriteLine("Page Count: " +
infoExtractor.GetPageCount());
        Console.WriteLine("Author: " +
infoExtractor.Author);
        Console.WriteLine("Title: " +
infoExtractor.Title);
        Console.WriteLine("Producer: " +
infoExtractor.Producer);
        Console.WriteLine("Subject: " +
infoExtractor.Subject);
        Console.WriteLine("CreationDate: " +
infoExtractor.CreationDate);
        Console.WriteLine("Text (first 2 lines): ");

        // Load a couple of lines from each document
        textExtractor.LoadDocumentFromFile(file);
        using (StringReader stringReader = new
StringReader(textExtractor.GetTextFromPage(0)))
        {
            Console.WriteLine(stringReader.ReadLine());
            Console.WriteLine(stringReader.ReadLine());
        }
        Console.WriteLine();
    }

    // Cleanup
    infoExtractor.Dispose();
    textExtractor.Dispose();

    Console.WriteLine();
    Console.WriteLine("Press any key to continue...");
    Console.ReadLine();
}
}
}
}
}

```

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](#)