

How to read PDF with scanned image and hindi text in C# and ByteScout Premium Suite

If you want to learn more then this tutorial will show how to read PDF with scanned image and hindi text in C#

The sample source codes on this page shows how to read PDF with scanned image and hindi text in C#. What is ByteScout Premium Suite? It is the set that includes 12 SDK products from ByteScout including tools and components for PDF, barcodes, spreadsheets, screen video recording. It can help you to read PDF with scanned image and hindi text in your C# application.

Want to quickly learn? This fast application programming interfaces of ByteScout Premium Suite for C# plus the guidelines and the code below will help you quickly learn how to read PDF with scanned image and hindi text. Simply copy and paste in your C# project or application you and then run your app! Further improvement of the code will make it more robust.

Our website gives trial version of ByteScout Premium Suite for free. It also includes documentation and source code samples.

FOR MORE INFORMATION AND FREE TRIAL:

[Download Free Trial SDK \(on-premise version\)](#)

[Read more about ByteScout Premium Suite](#)

[Explore API Documentation](#)

[Get Free Training for ByteScout Premium Suite](#)

[Get Free API key for Web API](#)

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

Source Code Files:

```
using System;
using System.Collections.Generic;
using System.IO;
using System.Text;
using System.Drawing.Imaging;
using Bytescout.PDFExtractor;
using Bytescout.PDF;
using System.Diagnostics;

namespace ReadPDFWithImageHindiText
{
    class Program
    {
        static void Main(string[] args)
        {
            try
            {
                // Files
                string fileName = "hindi_text_with_image.pdf";
                string destFileName = "output_hindi_text_with_image.pdf";
                string destFileName_serachable = "output_hindi_text_with_image_searchable.pdf";

                // Read all text from pdf file
                string allTextExtracted = "";
                using (TextExtractor extractor = new TextExtractor())
                {
                    // Load PDF document
                    extractor.LoadDocumentFromFile(fileName);

                    // Read all text directly
                    allTextExtracted = extractor.GetText();
                }

                // Get image from pdf file
                MemoryStream memoryStream = new MemoryStream();
                using (ImageExtractor extractor = new ImageExtractor())
                {
                    // Load PDF document
                    extractor.LoadDocumentFromFile(fileName);

                    if (extractor.GetFirstImage())
                    {
                        extractor.SaveCurrentImageToStream(memoryStream, ImageFormat.Png);
                    }
                }

                // Load image from file to System.Drawing.Image object (we need it to get image size)
                using (System.Drawing.Image sysImage = System.Drawing.Image.FromStream(memoryStream))
                {
                    // Compute image size in PDF units (Points)
                    float widthInPoints = sysImage.Width / sysImage.HorizontalResolution;
                    float heightInPoints = sysImage.Height / sysImage.VerticalResolution;

                    // Create new PDF document
                    using (Document outPdfDocument = new Document())
                    {
                        outPdfDocument.AddText(allTextExtracted);
                        outPdfDocument.AddImage(sysImage, widthInPoints, heightInPoints);
                    }
                }
            }
        }
    }
}
```

```

    {
        outPdfDocument.RegistrationName = "demo";
        outPdfDocument.RegistrationKey = "demo";

        // Create page of computed size
        Page page = new Page(widthInPoints, heightInPoints);

        // Add page to the document
        outPdfDocument.Pages.Add(page);

        Canvas canvas = page.Canvas;

        // Create Bytescout.PDF.Image object from loaded image
        Image pdfImage = new Image(sysImage);

        // Draw the image
        canvas.DrawImage(pdfImage, 0, 0, widthInPoints, heightInPoints);

        // Dispose the System.Drawing.Image object to free resources
        sysImage.Dispose();

        // Create brush
        SolidBrush transparentBrush = new SolidBrush(new ColorGray(0))

        // ... and make it transparent
        transparentBrush.Opacity = 0;

        // Draw text with transparent brush
        // Need to set Font which supports hindi characters.
        Font font16 = new Font("Arial Unicode MS", 16);
        canvas.DrawString(allTextExtracted, font16, transparentBrush, 4

        // Save document to file
        outPdfDocument.Save(destFileName);
    }
}

// Make PDF file with hindi text searchable to OCR.
using (SearchablePDFMaker searchablePDFMaker = new SearchablePDFMaker()
{
    //Load PDF document
    searchablePDFMaker.LoadDocumentFromFile(destFileName);

    // Set the location of "tessdata" folder containing language data
    /*
    * It used following files for hindi language support. Need to put
    * https://github.com/tesseract-ocr/tessdata/tree/3.04.00
    hin.traineddata
    hin.cube.bigrams
    hin.cube.lm
    hin.cube.nn
    hin.cube.params
    hin.cube.word-freq
    hin.tesseract_cube.nn
    */
    searchablePDFMaker.OCRLanguageDataFolder = @"c:\Program Files\Bytes

    // Set OCR language
    searchablePDFMaker.OCRLanguage = "hin";
}

```

```
// Need to set Font which supports hindi characters
searchablePDFMaker.LabelingFont = "Arial Unicode MS";

// Set PDF document rendering resolution
searchablePDFMaker.OCCRResolution = 300;

searchablePDFMaker.MakePDFSearchable(destFileName_serachable);
}

// Open document in default PDF viewer app
Process.Start(destFileName_serachable);
}
catch (Exception ex)
{
    Console.WriteLine("ERROR:" + ex.Message);
}

Console.ReadLine();
}
}
}
```

ReadPDFWithImageHindiText.sln

```
Microsoft Visual Studio Solution File, Format Version 12.00
# Visual Studio 15
VisualStudioVersion = 15.0.27703.2026
MinimumVisualStudioVersion = 10.0.40219.1
Project("{FAE04EC0-301F-11D3-BF4B-00C04F79EFBC}") = "ReadHindiText", "ReadPDFWithImageHindiText.sln", "{12345678-9ABC-DEF0-1234-56789ABCDEF0}"
EndProject
Global
    GlobalSection(SolutionConfigurationPlatforms) = preSolution
        Debug|Any CPU = Debug|Any CPU
        Release|Any CPU = Release|Any CPU
    EndGlobalSection
    GlobalSection(ProjectConfigurationPlatforms) = postSolution
        {99735776-2956-463D-9795-EBCE16928C30}.Debug|Any CPU.ActiveCfg = Debug|Any CPU
        {99735776-2956-463D-9795-EBCE16928C30}.Debug|Any CPU.Build.0 = Debug|Any CPU
        {99735776-2956-463D-9795-EBCE16928C30}.Release|Any CPU.ActiveCfg = Release|Any CPU
        {99735776-2956-463D-9795-EBCE16928C30}.Release|Any CPU.Build.0 = Release|Any CPU
    EndGlobalSection
    GlobalSection(SolutionProperties) = preSolution
        HideSolutionNode = FALSE
    EndGlobalSection
    GlobalSection(ExtensibilityGlobals) = postSolution
        SolutionGuid = {3ABE3EEF-B212-4E8B-9A74-67A52FD333AC}
    EndGlobalSection
EndGlobal
```



VIDEO

<https://www.youtube.com/watch?v=NEwNs2b9YN8>

ON-PREMISE OFFLINE SDK

[60 Day Free Trial](#) or [Visit ByteScout Premium Suite Home Page](#)
[Explore ByteScout Premium Suite Documentation](#)
[Explore Samples](#)
[Sign Up for ByteScout Premium Suite 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