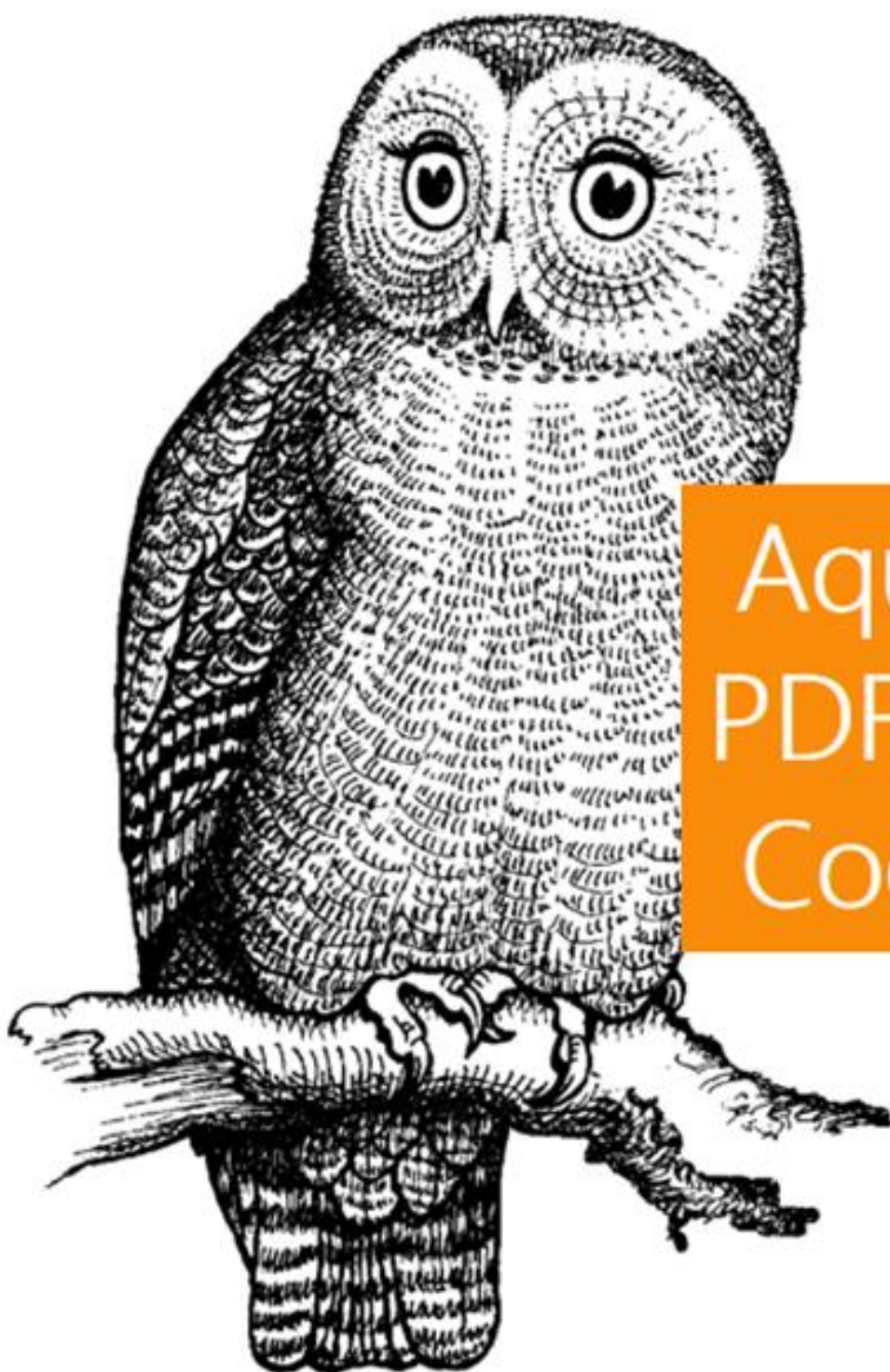


*PDF Solutions for C# and Microsoft.Net Developers*



# Aquaforest PDF Toolkit Cookbook

**Aquaforest**



Aquaforest  
PDF Toolkit

# Contents

<b>1</b>	<b>EXTRACT TEXT FROM PDF</b> .....	<b>1</b>
1.1	REQUIREMENT.....	1
1.2	SOLUTION.....	1
<b>2</b>	<b>USING THE PDF TOOLKIT WITH AQUAFORREST'S OCR SDK</b> .....	<b>2</b>
2.1	REQUIREMENT.....	2
2.2	SOLUTION.....	2
2.3	COMMENTS.....	4
<b>3</b>	<b>SPLIT PDF</b> .....	<b>5</b>
3.1	REQUIREMENT.....	5
3.2	SOLUTION.....	5
3.3	COMMENTS.....	5
<b>4</b>	<b>MERGE PDF</b> .....	<b>6</b>
4.1	REQUIREMENT.....	6
4.2	SOLUTION.....	6
4.3	COMMENTS.....	6
<b>5</b>	<b>CREATE A PDF DOCUMENT</b> .....	<b>7</b>
5.1	REQUIREMENT.....	7
5.2	SOLUTION.....	7
<b>6</b>	<b>APPLY STAMPS TO PDF FILES</b> .....	<b>8</b>
6.1	REQUIREMENT.....	8
6.2	SOLUTION.....	8
6.3	COMMENTS.....	8
<b>7</b>	<b>CONVERT AN IMAGE FILE TO PDF</b> .....	<b>9</b>
7.1	REQUIREMENT.....	9
7.2	SOLUTION.....	9
<b>8</b>	<b>EXTRACT IMAGES FROM PDF</b> .....	<b>10</b>
8.1	REQUIREMENT.....	10
8.2	SOLUTION.....	10
8.3	COMMENTS.....	10
<b>9</b>	<b>OVERLAY A PDF</b> .....	<b>11</b>
9.1	REQUIREMENT.....	11
9.2	SOLUTION.....	11
9.3	COMMENTS.....	11
<b>10</b>	<b>SET PDF SECURITY</b> .....	<b>12</b>
10.1	REQUIREMENT.....	12
10.2	SOLUTION.....	12
<b>11</b>	<b>GET AND SET XMP METADATA</b> .....	<b>13</b>

11.1	REQUIREMENT.....	13
11.2	SOLUTION.....	13
<b>12</b>	<b>CONVERT A CSV FILE TO PDF .....</b>	<b>14</b>
12.1	REQUIREMENT.....	14
12.2	SOLUTION.....	14

# 1 Extract Text from PDF

## 1.1 Requirement

Extract text from a PDF file.

## 1.2 Solution

```
using Aquaforest.PDF;
using System;

namespace ExtractTextFromPDF
{
    class Program
    {
        static void Main(string[] args)
        {
            PDFDocument doc = new
PDFDocument(@"..\..\..\documents\source\cookbook.pdf");

            string textFromWholeDocument = doc.GetText();

            string textFromPageFive = doc.GetText(5);

            Console.WriteLine(textFromWholeDocument);
            Console.WriteLine(textFromPageFive);
        }
    }
}
```

## 2 Using the PDF Toolkit with Aquaforest's OCR SDK

### 2.1 Requirement

Make use of the PDF Toolkit in conjunction Aquaforest's OCR SDK in order to:

- Check whether a file already has text
- Add a stamp on the output file
- Split or Merge results files

### 2.2 Solution

```
using Aquaforest.OCR.Api;
using Aquaforest.PDF;
using System;
using System.IO;
using System.Drawing;

namespace PDFToolkitWithOCRSDK
{
    class Program
    {
        static void Main(string[] args)
        {
            DirectoryInfo directory = new
DirectoryInfo(@"..\..\..\documents\source\tree");

            foreach (var file in directory.GetFiles("*.pdf",
SearchOption.AllDirectories))
            {
                Console.WriteLine("Processing PDF file: {0}",
file.FullName);
                PDFDocument doc = new PDFDocument(file.FullName);
                string docText = doc.GetText().Replace("\r\n",
"").Replace("\n", "").Replace("\r", "");
                int numberOfPages = doc.NumberOfPages;
                doc.Close();

                // OCR document only if it is image only
                if (docText.Length == 0)
                {
                    Console.WriteLine("PDF file is image-only");

                    string outputDirectory =
Path.Combine(@"..\..\..\documents\output", file.Directory.Name);
                    if (!Directory.Exists(outputDirectory))
                        Directory.CreateDirectory(outputDirectory);

                    string output = Path.Combine(outputDirectory,
file.Name);

                    bool ocrSuccessful = OCRPDF(file.FullName, output);

                    if (ocrSuccessful)
                    {
                        AddStamp(output);

                        if (numberOfPages > 1)
                        {
                            SplitFile(output);
                        }
                    }
                }
            }
        }
    }
}
```

```

        }
    }
    else
    {
        Console.WriteLine("PDF file contains text");
    }

    Console.WriteLine("");
}

static bool OCRPDF(string source, string output)
{
    bool success = false;

    try
    {
        Ocr ocr = new Ocr();
        PreProcessor preProcessor = new PreProcessor();

        string OCRFiles = Path.GetFullPath(@"..\..\lib\");
        Environment.SetEnvironmentVariable("PATH",
Environment.GetEnvironmentVariable("PATH") + ";" + OCRFiles);
        ocr.ResourceFolder = OCRFiles;
        ocr.EnableConsoleOutput = true;
        ocr.Language = SupportedLanguages.English;
        ocr.EnablePdfOutput = true;
        ocr.ReadPDFSource(source);

        preProcessor.Deskew = true;
        preProcessor.Autorotate = false;

        if (ocr.Recognize(preProcessor))
        {
            ocr.SavePDFOutput(output, true);
        }
        ocr.DeleteTemporaryFiles();

        success = true;
    }
    catch (Exception e)
    {
        Console.WriteLine("Error in OCR Processing : " + e.Message);
        success = false;
    }

    return success;
}

static void AddStamp(string output)
{
    Console.WriteLine("Adding stamp...");

    PDFDocument doc = new PDFDocument(output);

    PDFStamper stamper = new PDFStamper(doc, doc.FilePath);
    stamper.FontSize = 12;
    stamper.StampOpacity = 100;
    stamper.StampColor = Color.Black;
    stamper.StampPDFText("THIS IS A TEST STAMP", 200, 200);
}

```

```
static void SplitFile(string output)
{
    Console.WriteLine("Splitting file...");

    PDFDocument doc = new PDFDocument(output);

    PDFSplitter splitter = new PDFSplitter(doc);
    splitter.OutputFileName =
Path.GetFileNameWithoutExtension(output);
    splitter.OutputFilePath = Path.GetDirectoryName(output);
    splitter.SplitByRepeatingNumber(1, 5, 1);
}
}
```

### 2.3 Comments

The above code will loop through all the PDF documents in a directory and check if they either contain text or are image-only. The image-only PDFs are OCRed and stamped. If the OCRed documents have more than one page, they are split into single pages.

In order for this example to work you will need to download the OCR SDK (<http://www.aquaforest.com/en/ocrsdk.asp>) first and then reference Aquaforest.OCR.Api.

## 3 Split PDF

### 3.1 Requirement

Split PDF files.

### 3.2 Solution

```
using Aquaforest.PDF;

namespace SplitPDF
{
    class Program
    {
        static void Main(string[] args)
        {
            PDFDocument doc = new
PDFDocument(@"..\..\..\documents\source\cookbook.pdf");
            PDFSplitter splitter = new PDFSplitter(doc);
            splitter.OutputFileName = "split";
            splitter.OutputFilePath = @"..\..\..\documents\output\";

            splitter.SplitByRepeatingNumber(1, 5, 1);
        }
    }
}
```

### 3.3 Comments

The `PDFSplitter` class provides different methods to split PDF files namely:

- `SplitByRepeatingNumber`  
Split PDF document with each split containing a particular number of pages
- `SplitByTopLevelBookmarks`  
Split the document by the top level bookmarks in the PDF.
- `SplitByPageRanges`  
Split PDF document by page ranges. E.g. 1, 3-7, 10
- `SplitByRepeatingPageRanges`  
Apply the page range to each set of "Page Ranges" pages within the document. For example if 2-4 is specified for page ranges, and 4 is specified as the repeating range, then the range is re-applied every 4 pages. Hence the file is split such that the first output file contains pages 2-4 from the original document; the second contains pages 6-8 and so on.



## 4 Merge PDF

### 4.1 Requirement

Merge PDF files.

### 4.2 Solution

```
using Aquaforest.PDF;
using System.Collections.Generic;

namespace MergePDFs
{
    class Program
    {
        static void Main(string[] args)
        {
            PDFMerger merger = new PDFMerger();

            List<string> pdfList = new List<string>()
            {
                @"..\..\..\documents\source\image_pdf.pdf",
                @"..\..\..\documents\source\releasenotes.pdf",
                @"..\..\..\documents\source\pdf_with_images.pdf"
            };

            merger.MergePDFs(pdfList,
@"..\..\..\documents\output\merge.pdf");
        }
    }
}
```

### 4.3 Comments

The `PDFMerger` class provides 3 different methods of merging PDF files:

- Select and add individual PDF files to merge to a `List`
- Merge all PDF files in a particular folder
- Merge each folder in a folder tree separately

## 5 Create a PDF document

### 5.1 Requirement

Create a PDF document from scratch and add some text and metadata in it.

### 5.2 Solution

```
using Aquaforest.PDF;
using Aquaforest.PDF.Font;
using System;

namespace CreatePDF
{
    class Program
    {
        // Create a PDF document
        static void Main(string[] args)
        {
            PDFDocument doc = new PDFDocument();
            PDFPage page = new PDFPage();

            // Add text to page
            PDFPageContentStream contents = new PDFPageContentStream(doc,
page);

            contents.BeginText();
            contents.SetFont(PDFType1Font.COURIER_BOLD, 12);
            contents.MoveText(100, 700);
            contents.DrawString("Hello World!");
            contents.EndText();
            contents.Close();

            // Add metadata
            PDFDocumentInformation info = new PDFDocumentInformation()
            {
                Author = "Name Surname",
                Subject = "Test",
                Title = "New PDF",
                Keywords = "PDF, OCR, SDK",
                CreationDate = new DateTime(2013, 9, 9),
                Producer = "Aquaforest"
            };
            info.SetCustomMetadataValue("AQUAFOREST_PDF_TOOLKIT", "1.01");
            doc.SetDocumentInformation(info);

            // Add page to PDF document
            doc.AddPage(page);

            doc.Save(@"..\..\..\documents\output\pdf_out.pdf");
            doc.Close();
        }
    }
}
```

## 6 Apply Stamps to PDF Files

### 6.1 Requirement

Apply various types of stamp / watermark to PDF files.

### 6.2 Solution

```
using Aquaforest.PDF;
using System.Drawing;
namespace StampPDF
{
    class Program
    {
        static void Main(string[] args)
        {
            PDFDocument doc = new
PDFDocument(@"..\..\..\documents\source\releasenotes.pdf");

            PDFStamper stamper = new PDFStamper(doc,
@"..\..\..\documents\output\stamp.pdf");
            stamper.FontSize = 12;
            stamper.StampOpacity = 100;
            stamper.StampColor = Color.Black;
            stamper.StampPDFText("This is a test stamp", 200, 200);
        }
    }
}
```

### 6.3 Comments

The `PDFStamper` provides different types of stamping methods namely:

- `StampPDFText` - Stamp in the output PDF will be in image format
- `StampTextAsString` - Stamp in the output PDF will be in text format
- `StampPageNumber` - Add a number to each page starting from a particular number and incrementing by 1 for each page
- `StampPageNumberBates` - Stamp each page with a bates number
- `StampVariable` - Stamp each page with one of the stamp variables provided in the PDF Toolkit.

The table below shows all the acceptable variables:

Variable	Stamp
%a	Short Day (Mon)
%A	Long Day (Monday)
%b	Short Month (Jan)
%B	Long Month (January)
%c	Date and time (30 October 2013 17:21)
%C	Date and Time With seconds (30 October 2013 17:21:50)
%d	Month and Year (October 2013)
%D	Day and Month (30 October)
%e	Short Year (13)
%E	Long Year (2013)
%f	Short Time of Day (17:21)
%F	Time of Day With Seconds (17:21:20)
%G	Full Date and time (Wed, 30 October 2013 17:21:50 GMT)
%Y	File Name

## 7 Convert an Image file to PDF

### 7.1 Requirement

Convert an image file (BMP, PNG, JPEG, TIFF...) to PDF.

### 7.2 Solution

```
using Aquaforest.PDF;

namespace ImageToPDF
{
    class Program
    {
        static void Main(string[] args)
        {
            ImageToPDFConverter imageConverter = new
ImageToPDFConverter(@"..\..\..\documents\source\image.tif",
@"..\..\..\documents\output\output.pdf");
            imageConverter.ConvertImageToPDF();
        }
    }
}
```

## 8 Extract Images from PDF

### 8.1 Requirement

Extract all images from a PDF document.

### 8.2 Solution

```
using Aquaforest.PDF;

namespace ExtractImagesFromPDF
{
    class Program
    {
        static void Main(string[] args)
        {
            PDFDocument doc = new
PDFDocument(@"..\..\..\documents\source\pdf_with_images.pdf");
            PDFImageExtractor extractor = new PDFImageExtractor(doc);
            extractor.ExtractImages(@"..\..\..\documents\output", "img",
false);
        }
    }
}
```

### 8.3 Comments

Note that the images extracted from the PDF file will be saved in the same format that they were added to the PDF.

## 9 Overlay a PDF

### 9.1 Requirement

Add a PDF page as an overlay to another PDF.

### 9.2 Solution

```
using Aquaforest.PDF;

namespace AddOverlay
{
    class Program
    {
        static void Main(string[] args)
        {
            PDFDocument overlay = new
PDFDocument(@"..\..\..\documents\source\overlay.pdf");
            PDFDocument overlayDestination = new
PDFDocument(@"..\..\..\documents\source\image_pdf.pdf");
            PDFOverlay o = new PDFOverlay(overlay, overlayDestination);

            o.ApplyOverlay(@"..\..\..\documents\output\overlay_output.pdf");
        }
    }
}
```

### 9.3 Comments

Note that the page being added as an overlay should not be an image PDF or else it will cover the original page in the destination PDF completely.

## 10 Set PDF Security

### 10.1 Requirement

Set security and other access permissions to a PDF document.

### 10.2 Solution

```
using Aquaforest.PDF;

namespace SetPDFPermissions
{
    class Program
    {
        static void Main(string[] args)
        {
            PDFDocument doc = new
PDFDocument(@"..\..\..\documents\source\releasenotes.pdf");

            PDFPermission permission = new PDFPermission()
            {
                AllowExtractContents = false,
                AllowModifyContents = false,
                AllowAssembly = true,
                AllowDegradedPrinting = true,
                AllowExtractForAccessibility = true,
                AllowFillInForm = false,
                AllowModifyAnnotations = false,
                AllowPrinting = true
            };

            PDFSecurity encryptor = new PDFSecurity()
            {
                Permission = permission,
                OwnerPassword = "myOwnerPassword",
                UserPassword = "myUserPassword"
            };

            encryptor.SecurePDF(doc,
@"..\..\..\documents\output\secure_pdf.pdf");
        }
    }
}
```

## 11 Get and Set XMP metadata

### 11.1 Requirement

Get XMP metadata from a PDF file and set XMP metadata to a PDF file.

### 11.2 Solution

```
using Aquaforest.PDF;
using System;
using System.Xml;

namespace SetAndGetXMPMetadata
{
    class Program
    {
        static PDFDocument doc;

        static void Main(string[] args)
        {
            SetXMP();

            GetXMP();
        }

        static void SetXMP()
        {
            XmlDocument xmp = new XmlDocument();
            xmp.Load(@"..\..\..\documents\source\xmp.xml");

            doc = new
PDFDocument(@"..\..\..\documents\source\releasenotes.pdf");
            doc.SetXMPMetadata(xmp);
            doc.Save(@"..\..\..\documents\output\releasenotes_xmp.pdf");
        }

        static void GetXMP()
        {
            doc = new
PDFDocument(@"..\..\..\documents\source\pdf_with_images.pdf");
            string xmp = doc.GetXMPMetadata();

            Console.WriteLine(xmp);
        }
    }
}
```



## 12 Convert a CSV file to PDF

### 12.1 Requirement

Take a Comma Separated Value (CSV) file and convert it to a table in a PDF file.

### 12.2 Solution

```
using Aquaforest.PDF;
using Aquaforest.PDF.Font;

namespace CSVToPDF
{
    class Program
    {
        static void Main(string[] args)
        {
            CSVToPDFConverter csv = new
            CSVToPDFConverter(@"..\..\..\documents\source\csv_input.csv");

            // The page settings of the output PDF
            PDFPageSettings pageSettings = new PDFPageSettings();
            pageSettings.Font = PDFType1Font.HELVETICA;
            pageSettings.FontSize = 12;
            pageSettings.SetMargin(50, 50, 50, 50);
            pageSettings.SetSize(PageSize.A4);

            // The settings of the table in the output PDF
            PDFTableSettings tableSettings = new PDFTableSettings()
            {
                CellMargin = 5,
                RowHeight = 20
            };

            csv.ConvertToPDF(@"..\..\..\documents\output\csv_output.pdf",
            pageSettings, tableSettings);
        }
    }
}
```