# How to convert PDF to HTML C# sample

**Written by Apitron Documentation Team**

## Introduction

While PDF itself is a perfect format for storing documents, sometimes you might be interested in converting it to HTML for various reasons, e.g. for implementing quick presenter capable to provide text selection by using web browser control as a host, or for implementing PDF viewer for your website.

One of the problems is that PDF has very sophisticated drawing model, so it can't be translated one to one even using HTML5 features. It means that complex drawings won't be nicely converted using traditional methods, but we have implemented a solution that gives predictable results even for files containing advanced vector drawings.

Every graphical object will be converted to image using the same graphical engine as used in Apitron PDF Rasterizer product, while the text objects will be handled by the web browser. As a result you may expect PDF to HTML conversion to work fine in almost all cases.

## The code

The following code sample demonstrates how to convert PDF to HTML. As it can be seen from the code below, we're using *HtmlPage* option for this conversion with default resolution for images (72 dpi as PDF default). It's also possible to convert using *HtmlFragment* option creating a DIV object that can be embedded into the existing container (other DIV for example).

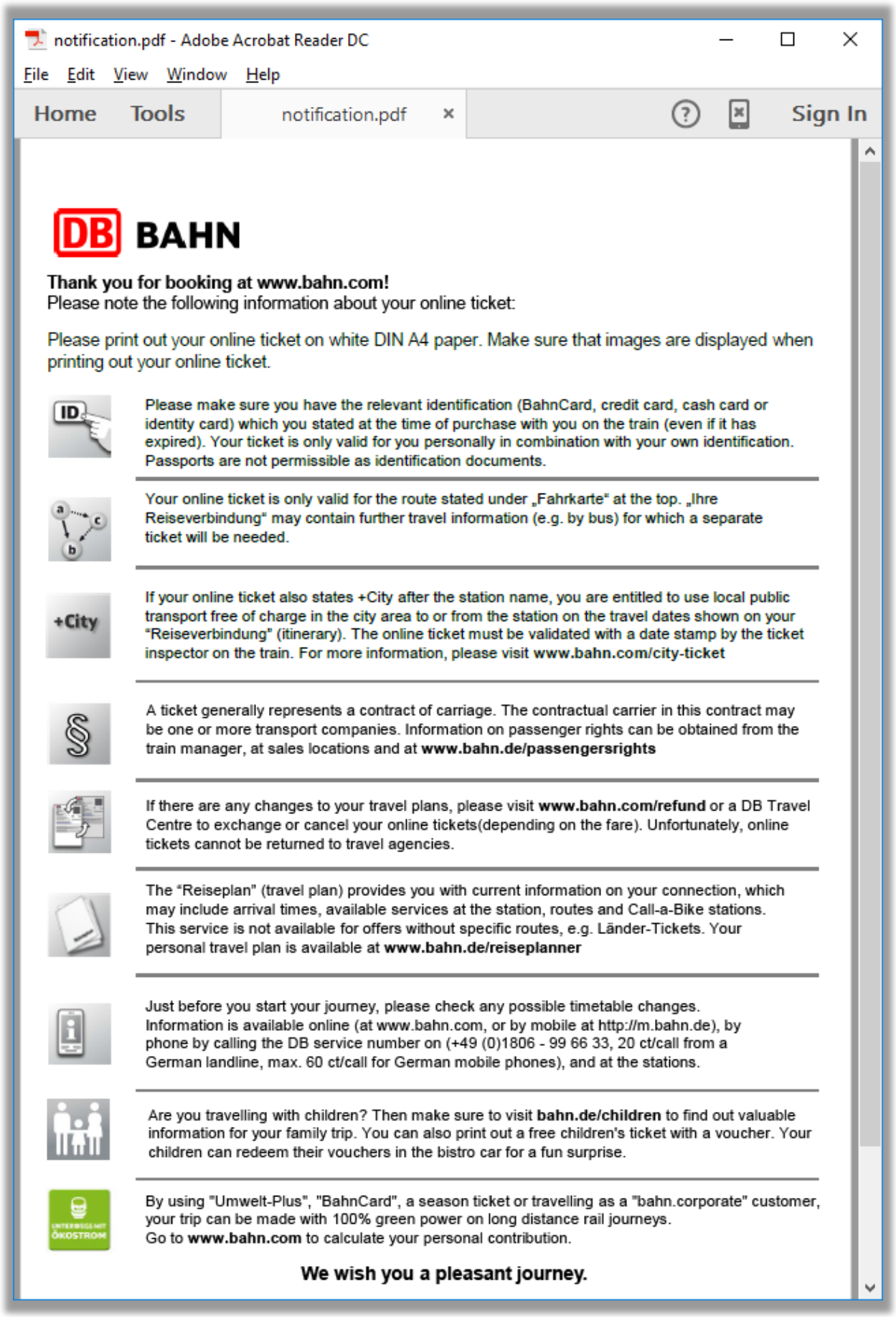
```
class Program
{
    static void Main(string[] args)
    {
        // open pdf document
        using (Stream inputStream = File.OpenRead("../../data/notification.pdf"))
        {
            using (FixedDocument doc = new FixedDocument(inputStream))
            {
                // create output file
                using (TextWriter writer =
                    new StreamWriter(File.Create("c:/out.html"),Encoding.UTF8))
                {
                    // write returned html string to file
                    writer.Write(doc.Pages[0].ConvertToHtml(TextExtractionOptions.HtmlPage));
                }
            }
        }

        Process.Start("c:/out.html");
    }
}
```

The complete example can be found in our github repository.

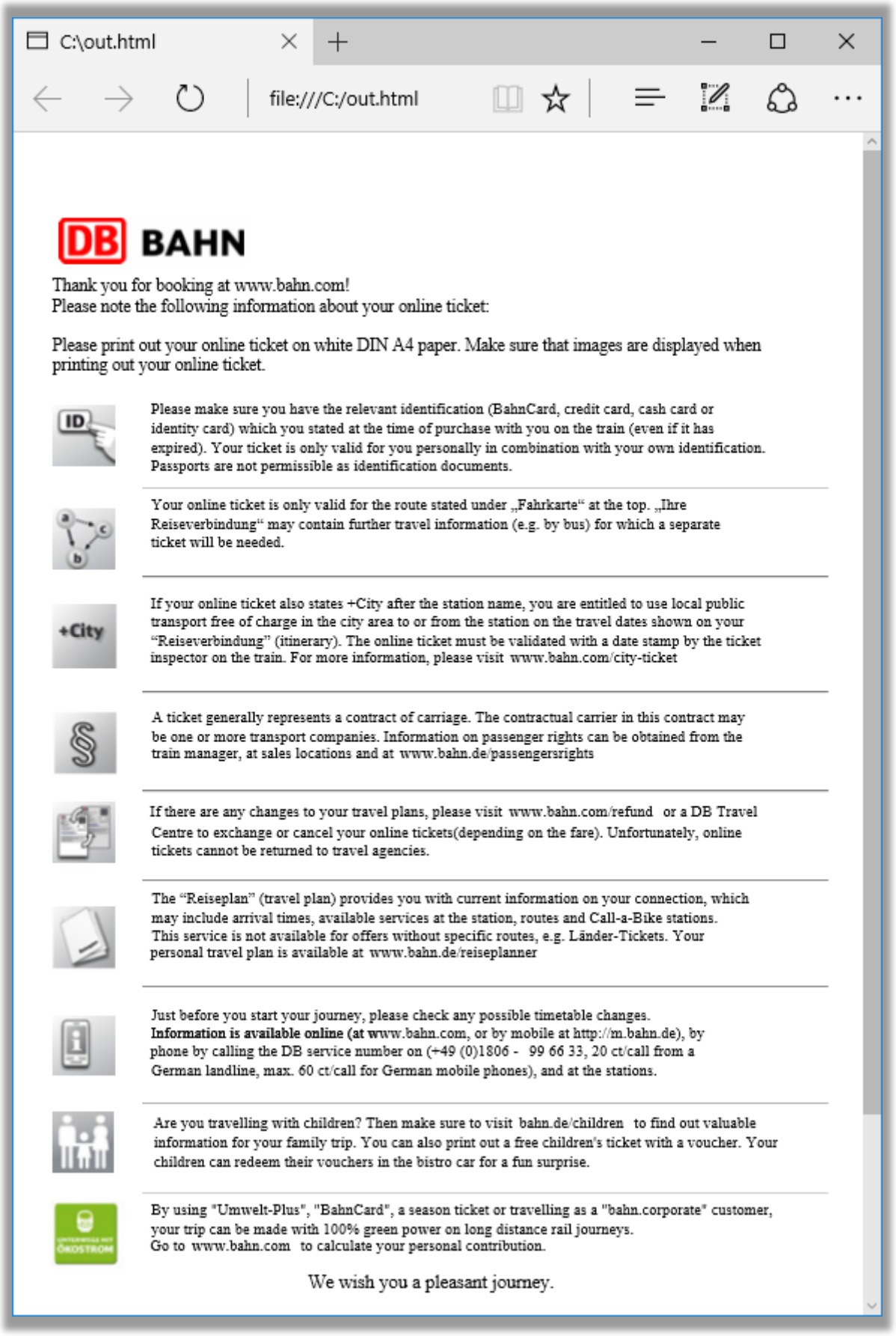
It's also possible to extract only drawings from PDF page, check page's *ExtractDrawings* method for the details.

Original PDF document opened in PDF viewer:




**Pic. 1 Original PDF file before conversion**

Converted from PDF to HTML using Apitron PDF Kit for .NET:




**Pic. 2 Converted HTML file opened in MS Edge**

## Summary

While PDF to HTML conversion is a hard task to implement it properly, we've made a significant improvement in this area and hope you'll like this new functionality added into the Apitron PDF Kit product. It's now in publicly available beta stage, so we'd highly appreciate your feedback and comments. Please don't hesitate to contact us with any questions or concerns.