Converting html to pdf using markup parser

Written by Apitron Documentation Team

Introduction

Sometimes, you have a pre-built HTML fragment and would like to convert it to PDF with minimal changes in its appearance. You could manually split it down to atomic elements like spans, DIVs et cetera, create corresponding pdf objects and get the job done. However, it may become a time consuming and boring task which could be easily automated using Apitron PDF Kit for .NET component. Strictly speaking, it's not an HTML to PDF converter in its traditional meaning, but it provides you with an API to parse any custom markup (provided it's XML-based) into pdf, and HTML is just the one of the possible markups.

Content wrapped by a tag will be converted to a flow layout content element with its `Class` property set to the tag's name. You can define appropriate styles, and style these elements using class selectors. Images can be added using reserved *<img>* tag. Attributes *link* and *bookmark* are reserved to produce internal or external links and bookmarks. *Id* attribute can be used to assign an Id to the element. While it all may sound a bit complicated it's actually pretty simple and convenient. Let's see the code sample from the next section.

If you'd like to read more about custom markup parsing, we have a free book for you - Apitron PDF Kit in action.

Basic HTML to PDF conversion

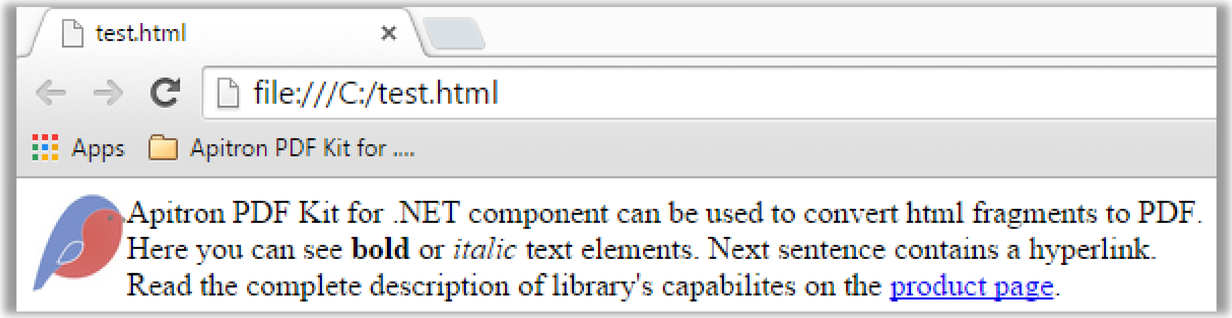
Source HTML

Consider the following html fragment:

```
...
<img src='myimage.jpg' height='48px' width='48px' style='float: left;'/>
Apitron PDF Kit for .NET component can be used to convert html fragments to PDF.
<br/>
Here you can see <b>bold</b> or <i>italic</i> text elements. Next sentence contains a hyperlink.
<br/>
Read the complete description of library's capabilities on the <a href='www.apitron.com/product/pdf-
kit'>product page</a>
...
```

It has bold and italic text, image element, and a link to other web page. All these elements will be converted to content elements defined in flow layout API subset and styled according to defined styles. You may also note the use of additional *<br>* element - it adds a line break.

In web browser it looks like this:



Apitron PDF Kit for .NET component can be used to convert html fragments to PDF.
Here you can see **bold** or *italic* text elements. Next sentence contains a hyperlink.
Read the complete description of library's capabilites on the product page.

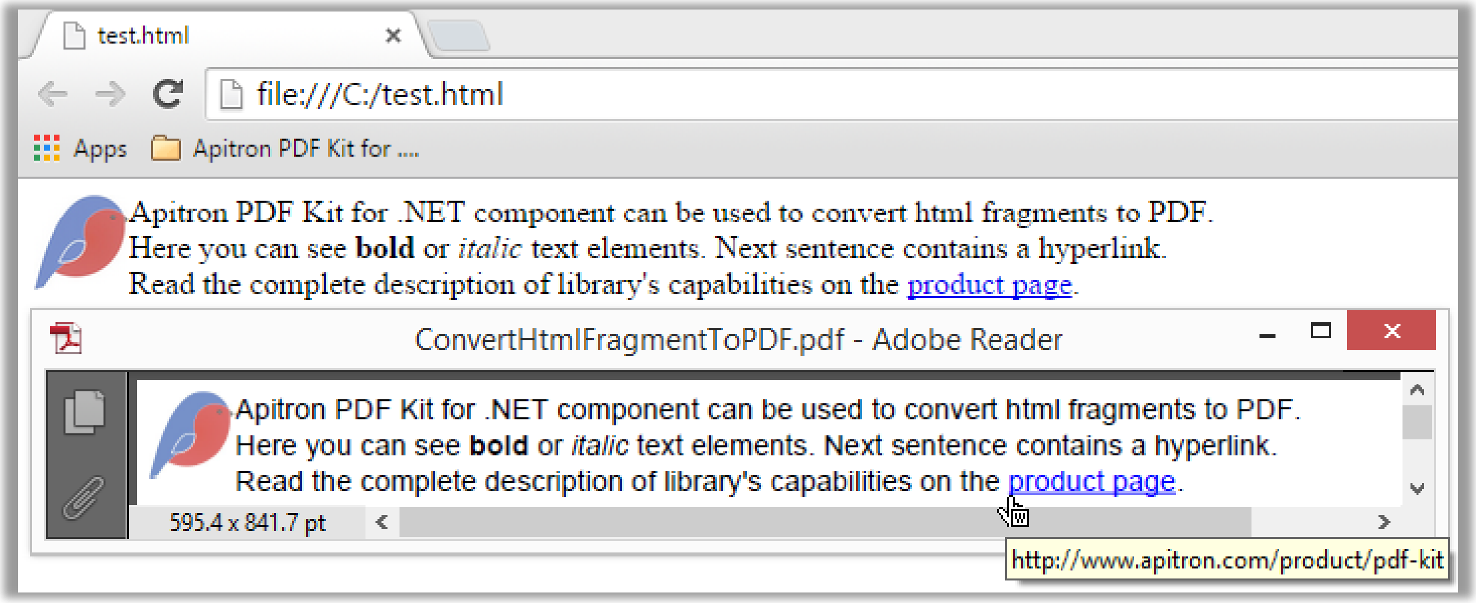
Pic. 1 HTML fragment sample

Conversion implementation

The following code performs the actual conversion:

```
public void ConvertHtmlFragmentToPDF()
{
    // prepare resources
    ResourceManager resourceManager = new ResourceManager();

    // create and fill document
    FlowDocument doc = new FlowDocument() { Margin = new Thickness(5, 5, 5, 5) };

    // register styles for different elements
    doc.StyleManager.RegisterStyle(".img", new Style(){Float = Float.Left});
    doc.StyleManager.RegisterStyle(".b",new Style(){Font=new Font(StandardFonts.HelveticaBold,12)});
    doc.StyleManager.RegisterStyle(".i",new Style(){Font=new Font(StandardFonts.HelveticaOblique,12)});
    doc.StyleManager.RegisterStyle(".a", new Style() { Color = RgbColors.Blue,
    TextDecoration = new TextDecoration(TextDecorationOptions.Underline)});

    // parse html markup
    doc.AddItems(ContentElement.FromMarkup("<img src='myimage.jpg' height='48px' width='48px'" +
    " style='float: left;'/> Apitron PDF Kit for .NET component can be used to convert html" +
    " fragments to PDF.<br/> Here you can see <b>bold</b> or <i>italic</i> text elements. " +
    "Next sentence contains a hyperlink.<br/>Read the complete description of library's" +
    " capabilities on the <a href='www.apitron.com/product/pdf-kit'>product page</a>."));

    // save document
    using (Stream stream = File.Create("ConvertHtmlFragmentToPDF.pdf"))
    {
        doc.Write(stream, resourceManager);
    }
}
```

Results of the conversion are as follows:


test.html

file:///C:/test.html

Apps Apitron PDF Kit for ....

Apitron PDF Kit for .NET component can be used to convert html fragments to PDF.
Here you can see **bold** or *italic* text elements. Next sentence contains a hyperlink.
Read the complete description of library's capabilities on the product page.

ConvertHtmlFragmentToPDF.pdf - Adobe Reader

Apitron PDF Kit for .NET component can be used to convert html fragments to PDF.
Here you can see **bold** or *italic* text elements. Next sentence contains a hyperlink.
Read the complete description of library's capabilities on the product page.

595.4 x 841.7 pt

http://www.apitron.com/product/pdf-kit

Pic. 2 Markup parsing results

Conclusion

This post demonstrates how to create basic html to pdf convertor and how to work with markup parsing API provided by Apitron PDF Kit. Create apps using your favorite .NET programming language: C#, VB .NET or any other. This library is available for many modern platforms, and you'll most likely find yours supported. Check out the samples included in download package and let us know if you have any questions or comments.