

XHTML TO PDF CONVERSION WITH XSL-FO*

Hannes Hirzel

This work is produced by The Connexions Project and licensed under the
Creative Commons Attribution License †

Abstract

The module shows an example how a subset of XHTML is converted with XSLT to XSL-FO (XSL Formatting Objects) and then to PDF.

NOTE: This module is a preliminary version.

1 Introduction

In this exercise an XHTML file using a small subset of tags is converted to a XSL-FO (XSL Formatting Objects) document¹ with an XSLT style² sheet. That file may be converted to PDF with a formatting objects processor³. In the reference section you find full size solutions.

To do the conversion you may use a setup with jEdit⁴.

The example XHTML input file is taken from this module⁵.

Sample XHTML input file to the XSLT transformation

Shortened version:

```
<?xml version="1.0" encoding="UTF-8"?>
<!DOCTYPE html>

<html>
<head>
<meta http-equiv="Content-Type" content="text/html; charset=UTF-8"/>
<title>De finibus bonorum et malorum</title>
</head>
<body>
<h1>De finibus bonorum et malorum</h1>

<h2>Sed ut perspiciatis</h2>
```

*Version 1.1: Jul 5, 2012 9:28 pm -0500

†<http://creativecommons.org/licenses/by/3.0/>

¹<http://cnx.org/content/m43592/latest/>

²<http://cnx.org/content/m43568/latest/>

³<http://cnx.org/content/m43600/latest/>

⁴<http://cnx.org/content/m43609/latest/>

⁵<http://cnx.org/content/m37275/latest/>

```

<p>Sed ut perspiciatis, unde omnis iste natus error sit voluptatem accusantium
doloremque laudantium, totam rem aperiam eaque ipsa, quae ab illo inventore veritatis
et quasi architecto beatae vitae dicta sunt, explicabo.
</p>

<p>Qui dolorem ipsum, quia <em>dolor sit amet</em>, consectetur
adipiscing velit, sed quia non numquam [do] eius modi tempora inci[di]dunt, ut
et dolore magnam aliquum quaerat voluptatem. Ut enim ad minima veniam, quis nostrum
exercitationem ullam corporis suscipit laboriosam, nisi ut aliquid ex ea commoda
consequatur?
</p>
    

<h2>At vero</h2>

<p>At vero eos et accusamus et <strong>iusto</strong> odio
dignissimos ducimus, qui blanditiis praesentium voluptatum deleniti atque corrup-
tos dolores et quas molestias excepturi sint, obcaecati cupiditate non provident,
similique sunt in culpa, qui officia deserunt mollitia animi, id est laborum et
fuga.
</p>
    
</html>

```

NOTE: The file uses the `` and `` tags. `` indicates emphasis and is often rendered as italic text whereas `` indicates stronger emphasis and is rendered as bold text (Ref.⁶)

XSLT style sheet

```

<?xml version="1.0"?>

<xsl:stylesheet xmlns:xsl="http://www.w3.org/1999/XSL/Transform"
    version="1.0"
    xmlns:fo="http://www.w3.org/1999/XSL/Format"
    xmlns:h="http://www.w3.org/1999/xhtml">
    <xsl:output method="xml"/>

    <xsl:template match="/">
        <fo:root>
            <fo:layout-master-set>

                <fo:simple-page-master master-name="simple"
                    page-height="29.7cm"
                    page-width="21cm"
                    margin-top="1cm"
                    margin-bottom="2cm"

```

⁶<http://www.w3.org/TR/html401/struct/text.html#h-9.2.1>

```
        margin-left="2.5cm"
        margin-right="2.5cm">
<fo:region-body margin-top="2.5cm"/>
<fo:region-before extent="2.5cm"/>
<fo:region-after extent="1.5cm"/>
</fo:simple-page-master>

</fo:layout-master-set>

<fo:page-sequence master-reference="simple">
    <fo:flow
        flow-name="xsl-region-body">
        <xsl:apply-templates />
    </fo:flow>
</fo:page-sequence>

</fo:root>
</xsl:template>

<xsl:template match="h:body">
    <fo:block id="{generate-id()}">
        <xsl:apply-templates/>
    </fo:block>
</xsl:template>

<xsl:template match="h:head">
    <fo:block
        font-family="Times"
        font-size="18pt"
        font-weight="bold"
        space-before="18pt"
        space-after="12pt"
        text-align="center">
        <xsl:apply-templates/>
    </fo:block>
</xsl:template>

<xsl:template match="h:p">
    <fo:block
        font-family="Times"
        font-size="12pt"
        space-before="12pt"
        space-after="12pt"
        text-align="justify">
        <xsl:apply-templates/>
    </fo:block>
</xsl:template>
```

```
<xsl:template match="h:em">
  <fo:inline
    color="red">
    <xsl:apply-templates/>
  </fo:inline>
</xsl:template>

<xsl:template match="h:strong">
<fo:inline font-weight="bold">
<xsl:apply-templates/>
</fo:inline>
</xsl:template>

<xsl:template match="h:h1">
  <fo:block
    font-family="Arial"
    font-weight="bold"
    font-size="14pt"
    space-before="12pt"
    space-after="12pt"
    text-align="justify">
    <xsl:apply-templates/>
  </fo:block>
</xsl:template>

<xsl:template match="h:h2">
<fo:block
  font-family="Arial"
  font-weight="bold"
  font-size="12pt"
  space-before="12pt"
  space-after="12pt"
  text-align="justify">
  <xsl:apply-templates/>
</fo:block>
</xsl:template>

<xsl:template match="h:img">
<fo:block text-align="center">
  <fo:external-graphic
    content-width="scale-to-fit"
    content-height="100%"
    width="100%"
    scaling="uniform"
    src="{@src}"/>
</fo:block>
</xsl:template>

</xsl:stylesheet>
```

The screenshot shows the JEdit interface with the following details:

- Title Bar:** JEdit - transform2fo.xsl
- Toolbar:** Standard JEdit toolbar with icons for file operations, search, and macros.
- Code Editor:** The main window displays the XSLT code for "transform2fo.xsl". The code defines a template for the root node, setting up a layout master set with a simple page master and specifying margins and regions.
- File Browser:** A sidebar on the right shows the file structure of "transform2fo.xsl", listing various XML elements like xsl:stylesheet, xsl:output, fo:root, etc.
- Status Bar:** Shows the file path (~/Documents/xhtml2fo/), encoding (UTF-8), file size (53/85 MB), and error count (1 error(s) 1:27 PM).

The result XSL-FO (Formatting Objects) document

```
<?xml version="1.0" encoding="UTF-8"?>
<fo:root xmlns:fo="http://www.w3.org/1999/XSL/Format" xmlns:h="http://www.w3.org/1999/xhtml">
<fo:layout-master-set>
<fo:simple-page-master margin-right="2.5cm" margin-left="2.5cm" margin-bottom="2cm" margin-top="1cm" pa
<fo:region-body column-count="2" margin-top="2.5cm"/>
<fo:region-before extent="2.5cm"/>
<fo:region-after extent="1.5cm"/>
</fo:simple-page-master>
</fo:layout-master-set>
<fo:page-sequence master-reference="simple">
<fo:static-content flow-name="xsl-region-before">
<fo:block line-height="14pt" font-family="serif" font-size="10pt" text-align="end">De finibus bonorum e
</fo:block>
</fo:static-content>
<fo:flow flow-name="xsl-region-body">

<fo:block margin-bottom="10mm" span="all" text-align="center" space-after="12pt" space-before="12pt" fo
```

```
<fo:block text-align="justify" space-after="12pt" space-before="12pt" font-size="12pt" font-weight="bol
```

```
<fo:block text-align="justify" space-after="12pt" space-before="12pt" font-size="12pt" font-family="Times New Roman">  
doloremque laudantium, totam rem aperiam eaque ipsa, quae ab illo inventore veritatis  
et quasi architecto beatae vitae dicta sunt, explicabo.  
</fo:block>  
  
<fo:block text-align="justify" space-after="12pt" space-before="12pt" font-size="12pt" font-family="Times New Roman">  
adipiscing velit, sed quia non numquam [do] eius modi tempora inci[di]dunt, ut  
et dolore magnam aliquum quaerat voluptatem. Ut enim ad minima veniam, quis nostrum  
exercitationem ullam corporis suscipit laboriosam, nisi ut aliquid ex ea commoda  
consequatur?  
</fo:block>  
  
<fo:block text-align="center">  
<fo:external-graphic src="aPicture.jpg" scaling="uniform" width="100%" content-height="100%" content-width="100%"/>  
</fo:block>  
  
<fo:block text-align="justify" space-after="12pt" space-before="12pt" font-size="12pt" font-weight="bold" font-family="Times New Roman">  
dignissimos ducimus, qui blanditiis praesentium voluptatum deleniti atque corrumptis,  
quos dolores et quas molestias excepturi sint, obcaecati cupiditate non provident,  
similique sunt in culpa, qui officia deserunt mollitia animi, id est laborum et fuga.  
</fo:block>  
  
<fo:block text-align="justify" space-after="12pt" space-before="12pt" font-size="12pt" font-family="Times New Roman">  
<fo:block text-align="center">  
<fo:external-graphic src="863px-Partridge_PSF.png" scaling="uniform" width="100%" content-height="100%"/>  
</fo:block>
```

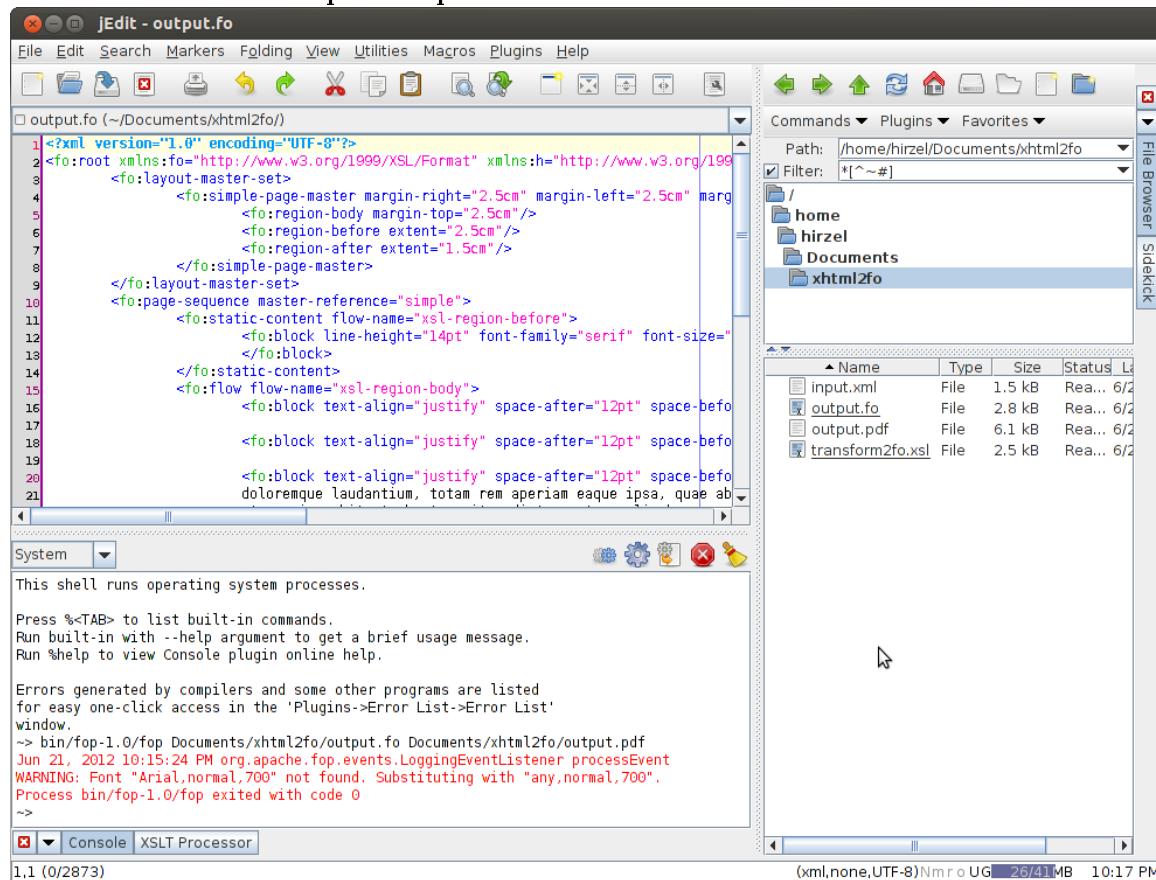
```
</fo:block>
```

```
<fo:block text-align="justify" space-after="12pt" space-before="12pt" font-size="12pt" font-family="Tim
</fo:block>

<fo:block text-align="justify" space-after="12pt" space-before="12pt" font-size="12pt" font-family="Tim
</fo:block>
```

```
</fo:flow>
</fo:page-sequence>
</fo:root>
```

XSL-FO to PDF with Apache fop



2 Download

input.xml⁷

⁷ <http://cnx.org/content/m43711/1.1/input.xml>

transform2fo.xsl⁸
 output.fo⁹
 output.pdf¹⁰

3 References

HTML to Formatting Objects (FO) conversion guide¹¹ (IBM developer works)

The code from IBM developer works has been adapted as an example which comes with Apache fop. Folder **x-xslfo2app-samples** in the folder **fop**. It contains the files

- everything.html
- xhtml-to-xslfo.xsl
- everything.fo
- everything.pdf

There is as well a shorter version

- short.html
- xhtml-to-xslfo-short.xsl

Comprehensive solution by XMLmind http://www.xmlmind.com/xhtml5_resources.shtml¹² ; XHTML5 schema plus XSLT for conversion to XSL-FO.

Another project, <http://xhtml2fo-xsl.googlecode.com/>¹³ , the XSLT file¹⁴

⁸<http://cnx.org/content/m43711/1.1/transform2fo.xsl>

⁹<http://cnx.org/content/m43711/1.1/output.fo>

¹⁰<http://cnx.org/content/m43711/1.1/output.pdf>

¹¹<http://www.ibm.com/developerworks/library/x-xslfo2app/>

¹²http://www.xmlmind.com/xhtml5_resources.shtml

¹³<http://xhtml2fo-xsl.googlecode.com/>

¹⁴<http://xhtml2fo-xsl.googlecode.com/svn/trunk/xhtml2fo.xsl>