HTML2fo - a HTML to XSL:FO converter

HTML2FO ist a converter for HTML files to the new XSL:FO format. It supports most of the usual tags. If you are missing a tag or think a tag is not handled as expected please open a <u>feature request</u> item. You may think that you have a XSLT which does the same job. But html2fo does convert documents which are **not** XML conform.

<u>Here is my Project site</u> at

You may <u>lock down</u> for a example pdf file.

Origin

I have developed html2fo because I had to create a new server driven printing solution for an client-server-based application. The previous printing solution was using Microsoft Word mailing function for importing a csv like text file and printing. As everybody knows - Word is not platform independent. But this was the main goal for the new printing solution. We have chosen PDF as platform independent document format and I had to convert about 40 documents with about 100 Sheets altogether. I used StarOffice to convert from .doc to .html because Word is in HTML export not as good as StarOffice. (There are worlds between them...) After using html2fo for converting to xsl:fo, a manual processing and rendering to PDF using FOP from Apache - Now I have a new printing solution.

html2fo supports:

- non-well-formed HTML-code
  the code will not be correct processed but you will get an output. This is good if you are using a bad WYSIWYM-editor like Word for editing HTML-files...
  This does not work at all. If it is too bad you will get a core dump... ;-)

- tables

  - colspans
    with a automatic column-width setting.
    If a non-"colspan"ed cell has a width setting - the corresponding column gets the width. Within the second run I am trying to calculate the width from col-spanned cells. The remaining space is divided through the rest of columns - this will happen for tables without a column with information

  - rowspans
    are completly supported - also in combination with colspans

  - Borders
    due non supported cell borders in HTML you could decide whether every cell has a border or none.

  - background color

- Font information:

  - Size

  - Style like Bold, Italic, Underline

  - Color

- Links

  both internal and external links are supported. A combination like referrered_file.html#marker is converted to a external reference.

  A reference to a .htm or .html file is converted to .pdf except the basename is the same as the converted file.

Architecture

html2fo converts the commons with an simple internal table and converts complex differences within functions. By using this way it is very simple to add a new HTML tag or Property.

Downloads

official releases

CVS web interface

html2fo

Mailing lists

announce
devel
users

Links

html2fo - html to xsl:fo (my project site at SourceForge)
FOP from Apache - xsl:fo to PDF (it's free) (you may look to the example section below)
XEP from RenderX - xsl:fo to PDF (it's not free)
jfor - xsl:fo to RTF (it's free but incomplete, not stable and has currently a confusing output)
Extensible Stylesheet Language (XSL) from W3C (also available as converted PDF)

PDF Examples

Every PDF file is rendered using FOP.
Every RTF file is rendered using JFOR.

This file as PDF or as RTF is only an example. Here is the file in the middle(XSL:FO).

My Test Suite

badformed.html (code)	badformed.fo	badformed.pdf
img.html (code)	img.fo	img.pdf
table.html (code)	table.fo	table.pdf

The complete FOP homepage as crosslinked PDF files is available here

The Proposed Recommendation of XSL:FO specification (267 tables, 47 images) as PDF (336 pages, 2.5MB) or as RTF (~ 272 pages, 5.3 MB).