



FOSDEM'14

Improving the XHTML export filter

And mentoring students
through patches and licenses

Andrea Pescetti
pescetti@apache.org

An Obscure Little Feature



An Interesting Story

- A little feature
- Students
- Forks
- Licenses

The XHTML Export

- One of many export filters
- File → Export → XHTML
- Not the HTML export!

Exporting to XHTML

- Available in all modules, but focusing on **Impress** here.
- Produces one `.html` file, not a slide-by-slide gallery.
- Code: `main/filter/source/xslt`

Habib's Work

- Improvements by Habib Louafi, a PhD student in Canada.
- Extends the XHTML filter with features that make it mature.
- Published in May 2013 (Louafi-Coulombe).

Habib's Code

- Full diff: 6666 lines.
- Based on versions 2.4 (2008) and 3.1 (2009).
- Contributed under Apache License v2 in June 2013:
[#i122494#](#)



New Features

Improving Fonts

- Fonts used to be replaced by the browser's default font.
- `style_collector.xsl` collects styles, `style_mapping_css.xsl` converts them to CSS.
- Now fixed: font is preserved.

Improving Layout

- Text boxes in a slide used to be exported sequentially in XHTML.
- All layout lost (body.xml)
- Now properly placed with coordinates and CSS.

Improving Images

- Images used to be converted to base64 and rendered as data streams embedded in XHTML.
- Problems with some browsers.
- Now solved: the “Link” option is used to preserve image.

Improving Shapes

- Shapes used to be ignored.
- Now converted using the **JavaScript Vector Graphics Library** by Walter Zorn.
- Support for rectangles, circles, ellipses, lines.

Improving Scaling

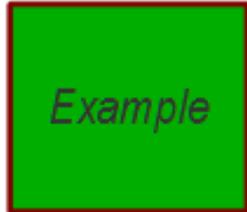
- The export did not support scaling.
- Scaling parameter added to resize the output (for mobile).
- Done by calling ImageMagick from a Java-based application.

Before and After (by Habib)

- Fonts
- Layout
- Images
- Shapes
- Scaling

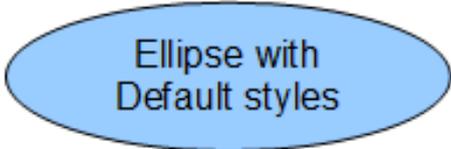
Default font

This paragraph contains four lines
This is the second line
This is the third line
AND THIS THE LAST LINE OF THE PARAGRAPH



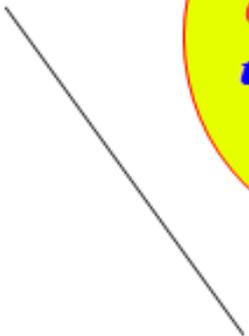
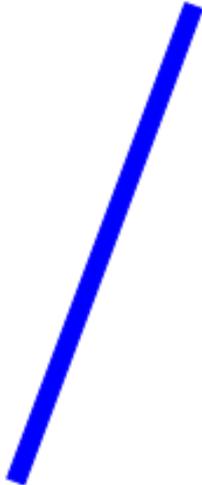
Times New Roman

Arial – not default font



Book Antiqua

Courrier New



Rectangle with default styles

Text inside a rectangle
Another text

Circle two-1

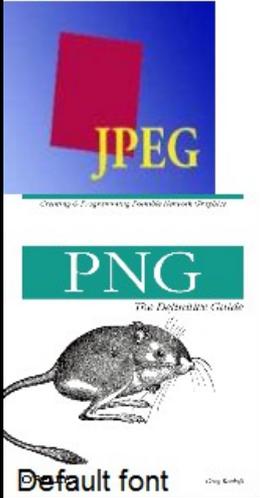
two-2

Ellipse Three-1
Three- 2

Text inside a rectangle

Another text

Example



Rectangle with default styles

Times New

Roman not default font

Book Antiqua

Courier New

This paragraph contains four lines

This is the second line

This is the third line

And this the last line of the paragraph

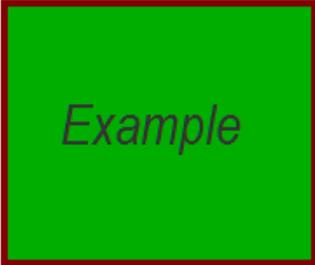
Default font

This paragraph contains four lines

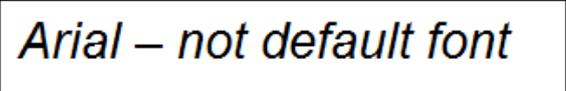
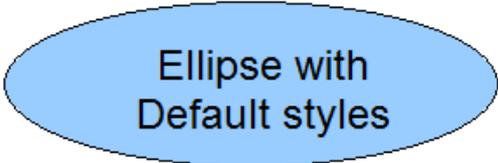
This is the second line

This is the third line

AND THIS THE LAST LINE OF THE PARAGRAPH

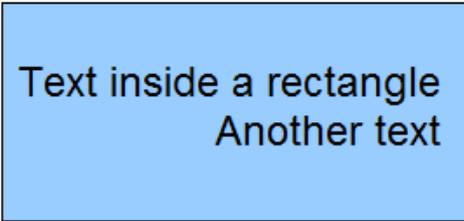
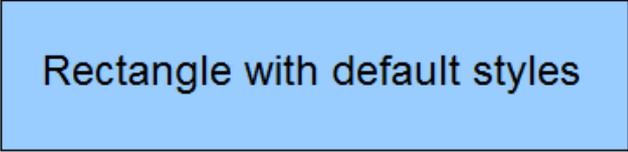
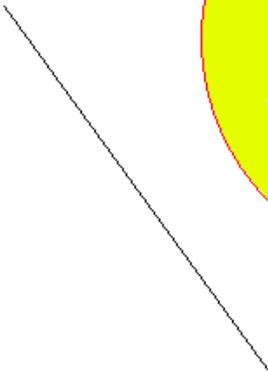
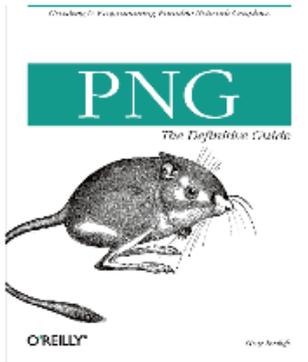
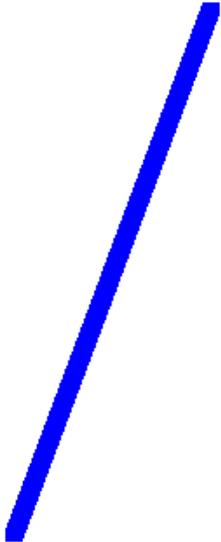


Times New Roman



Book Antiqua

Courrier New



Integration and difficulties



Four Years Are Long

- Habib's code cannot be used immediately.
- Code in both OpenOffice and LibreOffice has evolved, entire directories were reorganized.
- Need to refactor the patch.

Where's the XHTML Export?

- On Windows, XHTML export is not installed by default.
- Decision taken (undocumented) before 3.3.0: [#i119663#](#)
- Need to enable it with a custom installation.

Dependency Issues

- Habib's approach includes in full the late (to the best of our knowledge) Walter Zorn's JavaScript library.
- LGPL 2.1+, Apache Category B.
- Good case for Apache Legal.

Current Status

- Vadim Yedzinovich, a new contributor, is being mentored to work on this.
- Patch aligned and being sliced in several components.
- First fixes to appear in 4.1.

Future Developments

- Repurpose the XHTML export to an HTML5 export.
- Scaling: CSS3's transform.
- Shapes: HTML5's `<canvas>` and no JavaScript.
- Volunteers welcome!

Thanks for attention

Andrea Pescetti

pescetti@apache.org

www.openoffice.org

Image credits: Flickr, PLI0 Archives.