

Different “Hello, World!” PDF documents (generated by applications)

The directory *different-hello-worlds-generated* contains a few simple ‘Hello, World’ documents, created with the help of different PDF producing software applications.

We attempted to use the same fonts and same text positions on the page.

The files’ purpose is to provide some real world examples about how different tools may create very different PDF source code even if the page appearance is almost identical.

Comparing page appearances

You can use ImageMagick’s `compare` command (and variations) to create “delta” PDFs between different versions of the file:

```
compare          hw-ghostscript910.pdf hw-libreoffice44.pdf      delta1.pdf
compare -density 200 hw-ghostscript916.pdf hw-googledocs.pdf      delta2.pdf
compare          hw-ghostscript916.pdf hw-goggleslides.pdf -compose src delta3.pdf
compare -fuzz 5%  hw-googledocs.pdf    hw-goggleslides.pdf -compose src delta4.pdf
```

Note: When rendering PDFs, `compare` employs Ghostscript as its “*delegate*”, because ImageMagick cannot handle PDF input all by itself. The resulting “delta” files will show red pixels where there are differences on the pages. (`-density` controls the resolution used to render the pages. `-compose src` creates white+red pages only – skipping it will use a pale version of the first file as a background image. `-fuzz` will set a threshold below which smaller differences in colors are ignored.)

Can you point out why some of the “delta” documents are predominantly red, while others deviate only in smaller number of pixels?

Comparing PDF file structures and contents

You should use all tools you learn about to explore the differences amongst these files. Some suggestions:

- `pdftinfo` and `pdftinfo -meta` :
- `pdffonts` and `pdffonts -subst` :
- `pdfid.py` :
- `pdf-parser.py` :
- `peepdf.py` :
- `Vim` :

Note: `pdftinfo` and `pdffonts` do not support some of the suggested options if you use the venerable XPDF versions. A more recent release of the Poppler fork is required.