Fonts and Solid Framework

Author: Roger Dunham

Date: 6th March 2019

Introduction

When a document is reconstructed from a PDF, an attempt it made to use the same font in the document as was used in the PDF.

There are times when this is not possible.

For example, if the PDF uses an obscure font (see Figure 1), then Solid Framework will need to substitute the original font for the one that it considers to be most similar. Other scenarios where substitution commonly occurs are when the PDF contains "Times" or other fonts which is generally not available on a Windows machine and must therefore be substituted with "Times New Roman".

Figure 1- Example of an unusual font

Q: Why can't Solid Framework just extract the font from the original PDF?

A: Very often the PDF only contains the subset of the font for the characters that are actually used. That is very efficient in the PDF, but not very useful if you need to use a different character in the reconstructed document.

Furthermore, many fonts are subject to licenses which prevent them being copied from the original PDF.

In previous versions of Solid Framework, the fonts that were considered as possible substitutes were just those that were installed on the machine on which Solid Framework was running. This worked well on Windows, but Linux generally has far fewer and different fonts to those available on Windows machines. Some versions of Windows Server also have very limited fonts available.

As such if a document is reconstructed on a Linux server then the result may be very different from the result that would be obtained by converting on a Windows desktop machine. Furthermore, there is no guarantee that the conversion would be the same on two different windows desktop machines, since different fonts could be installed on each.

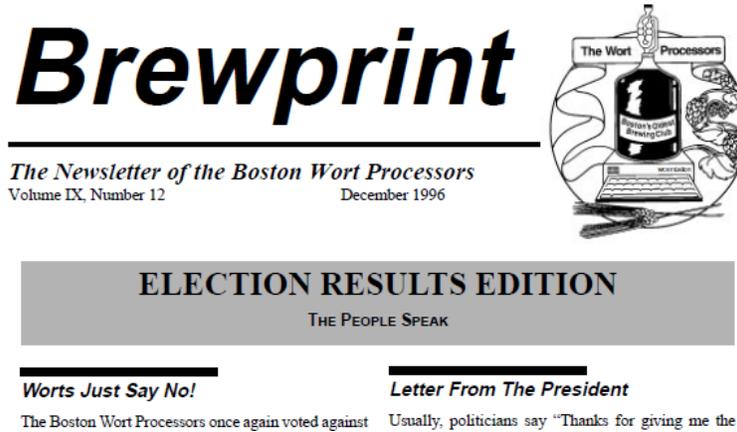The problem of differing reconstruction is shown below:
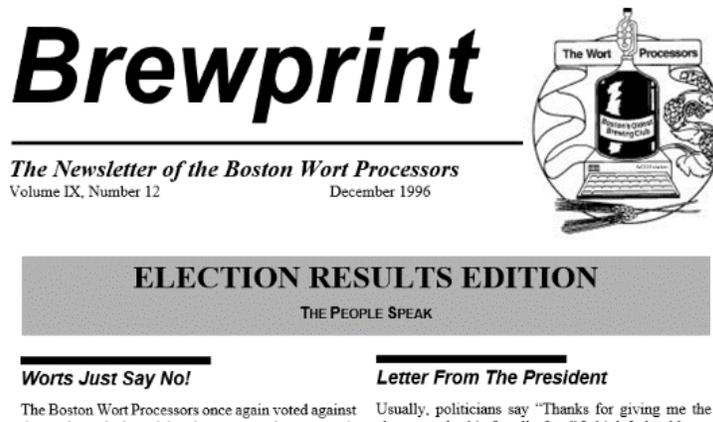

Figure 2 - Original PDF

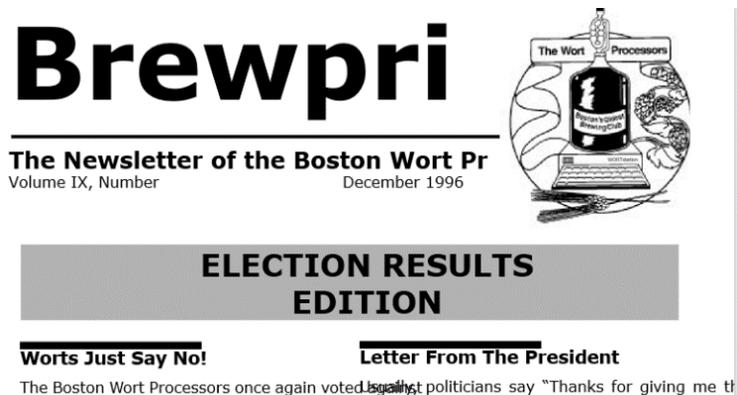Figure 3 - Document reconstructed on a Windows desktop machine.

Figure 4 Document reconstructed on a Linux machine with limited fonts.

The main issue here is that the Arial font in the original PDF has been substituted for DejaVu Sans, since Arial was not available

A secondary issue is that when viewed on a Windows machine, Word has further substitued the font DejaVu Sans with Verdana.

Specifying an Alternative Folder to Search for Fonts

Although the default location that is searched for font information is the system's fonts folder, it is possible to specify an alternative location where Solid Framework will search for fonts. This can be done using:

```
SolidFramework.Platform.Platform.SetFontDirectory([Path to folder containing fonts])
```

However this still requires the font files to be physically located within the specified directory.
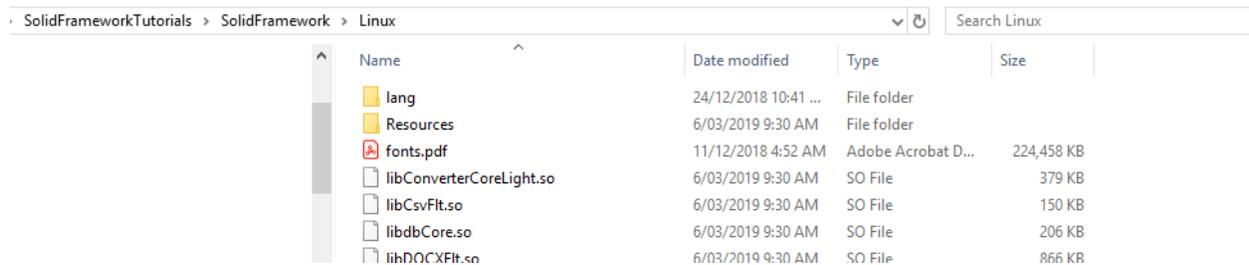
Using a Fonts Database File

Solid Framework 9182 introduced the concept of a "fonts" database file. This file is normally called "fonts.pdf". Solid Documents have created a version of this file that contains the most commonly used fonts. This is available for download, but is not shipped as part of Solid Framework due to its size (more than 200MB).

> Q: What if I want a customised set of fonts?
>
> A: It is possible to create a customised version of this file to include or exclude fonts based on your specific requirements. Please email support@soliddocuments.com if you require further information about how to do this.

Solid Framework will, by default, search for the fonts.pdf in the folder[1] that contains the unpacked Solid Framework library files (see figure 5). If the file is found, then Solid Framework will *only use the fonts embedded in this file* for substitution.

Figure 5- Typical location of fonts.pdf - in this case within a Linux installation.

Note that if the fonts.pdf file is found, then fonts that are installed on the machine, but which are not present in the fonts.pdf file will not be used in the reconstructed document.

---

[1] Alternatively the file can be located at an arbitray location with this being specified using

SolidFramework.Platform.Platform.SetFontsDataBaseFile([*Full path to fonts file])*

Using the fonts file can result in significantly improved reconstruction even on a Linux machine with few installed fonts.

Figure 6 - Document reconstructed on a Linux machine using fonts.pdf

Summary and Recommendation

If working on a platform which has limited fonts available, or where consistent reconstruction is required regardless of the machine on which conversion occurs, then it is recommended to use the downloadable "fonts.pdf" file that contains the common fonts required for document reconstruction.