

extractpdfmark

Extract page mode and named destinations as PDFmark from PDF

<https://github.com/trueroad/extractpdfmark>

<http://www.ctan.org/pkg/extractpdfmark>

When you create a PDF document using something like a TeX system you may include many small PDF files in the main PDF file. It is common for each of the small PDF files to use the same fonts.

If the small PDF files contain embedded font subsets, the TeX system includes them as-is in the main PDF. As a result, several subsets of the same font are embedded in the main PDF. It is not possible to remove the duplicates since they are different subsets. This vastly increases the size of the main PDF file.

On the other hand, if the small PDF files contain embedded full font sets, the TeX system also includes all of them in the main PDF. This time, the main PDF contains duplicates of the same full sets of fonts. Therefore, Ghostscript can remove the duplicates. This may considerably reduce the main PDF-file's size. (Note: Ghostscript 9.17 - 9.21 needs

-dPDFDontUseFontObjectNum

commandline option for removing duplicate fonts. If you use Ghostscript 9.22+, you cannot use this "full set embedding" method since it cannot remove duplicate fonts. In this case, you can use "not embedding" method as following.)

Finally, if the small PDF files contain some fonts that are **not** embedded, the TeX system outputs the main PDF file with some fonts missing. In this case, Ghostscript can embed the necessary fonts. It can significantly reduce the required disk size.

Either way, when Ghostscript reads the main PDF produced by the TeX system and outputs the final PDF it does not preserve PDF page-mode and named-destinations etc. As a result, when you open the final PDF, it is not displayed correctly. Also, remote PDF links will not work correctly.

http://bugs.ghostscript.com/show_bug.cgi?id=696943

http://bugs.ghostscript.com/show_bug.cgi?id=695760

This program is able to extract page mode and named destinations as PDFmark from PDF. By using this you can get the small PDF files that have preserved them.

extractpdfmark

Usage:

```
extractpdfmark TeX-Output.pdf > PDFmark.ps  
rungs -q -dBATCH -dNOPAUSE -sDEVICE=pdfwrite \  
-sOutputFile=Final.pdf TeX-Output.pdf PDFmark.ps
```

Git repository:

<https://github.com/trueroad/extractpdfmark>

Licence:

Copyright (C) 2016-2019 Masamichi Hosoda

Extract PDFmark is free software: you can redistribute it and/or modify it under the terms of the GNU General Public License as published by the Free Software Foundation, either version 3 of the License, or (at your option) any later version.

Extract PDFmark is distributed in the hope that it will be useful, but WITHOUT ANY WARRANTY; without even the implied warranty of MERCHANTABILITY or FITNESS FOR A PARTICULAR PURPOSE. See the GNU General Public License for more details.

You should have received a copy of the GNU General Public License along with Extract PDFmark.

If not, see <<http://www.gnu.org/licenses/>>.