pdfcpu

A PDF processor written in Go.

View the Project on GitHub pdfcpu/pdfcpu

About
Getting Started
Fonts
Core
- Collect
- Crop
- Merge
- **Optimize**
- Resize
- Rotate
- Split
- Stamp
- Trim
- Validate
- Watermark

Generate
Annotations
Attachments
Boxes
Create
Encryption
Extraction
Forms
Images
Info
Keywords
Pages
Papersizes
Portfolio
Properties

Changelog

Optimize

Optimize `inFile` by getting rid of redundant page resources like embedded fonts and images and write the result to `outFile` maxing out PDF compression. Have a look at some examples.

Usage

```
pdfcpu optimize [-stats csvFile] inFile [outFile]
```

Flags

name	description	required
stats	CSV output file	no

Common Flags

name	description	values
v(erbose)	turn on logging	
vv	verbose logging	
q(uiet)	quiet mode	
u(nit)	display unit	po(ints),in(ches),cm,mm
c(onf)	config dir	$path, disable
upw	user password	
opw	owner password	

Arguments

name	description	required	default

name	description	required	default
inFile	PDF input file	yes	
outFile	PDF output file	no	inFile

Stats

The name of a CSV file name.
This command appends one CSV line with stats about memory usage, PDF object usage and other useful information for debugging. Optimize a group of PDF input files and consolidate stats into the same CSV file for comparison.

The following shows a stats file with its header line and a single stats line:

```
cat stats.csv
name;version;author;creator;producer;src_size (bin|text);src_
test.pdf;1.2;;;;6 KB (67.4% | 32.6%); 0.0% |  0.0% | 100.0%;5
```

Examples

Optimize `test.pdf` and write the result to `test_new.pdf`:

```
pdfcpu optimize test.pdf
writing test_new.pdf ...
```

Optimize `test.pdf` and write the result to `test_opt.pdf`:

```
pdfcpu optimize test.pdf test_opt.pdf
writing test_opt.pdf ...
```

Optimize `test.pdf`, write the result to `test_opt.pdf`, append stats to `stats.csv` and produce logging on standard out:

```
pdfcpu optimize -verbose -stats stats.csv test.pdf test_opt.p
```

```
stats will be appended to stats.csv
 INFO: 2019/02/20 23:20:12 reading upc.pdf..
 INFO: 2019/02/20 23:20:12 PDF Version 1.5 conforming reader
 INFO: 2019/02/20 23:20:12 validating
 INFO: 2019/02/20 23:20:12 optimizing fonts & images
STATS: 2019/02/20 23:20:12 XRefTable:
*************************************************************
HeaderVersion: 1.2
has 2 pages
XRefTable:
                     Size: 13
              Root object: (11 0 R)
              Info object: (12 0 R)
                ID object: [<81C4A57DF6A1E411BD62885083B053CD
XRefTable with 13 entres:
    0: f   next=       0 generation=65535
    1:   offset=      16 generation=0 pdfcpu.Dict type=Page
<<
        <Contents, (2 0 R)>
        <Parent, (3 0 R)>
        <Resources, (4 0 R)>
        <Type, Page>
>>
    2:   offset=     102 generation=0 pdfcpu.StreamDict
<<
        <Filter, LZWDecode>
        <Length, 2652>
>>
    3:   offset=    5117 generation=0 pdfcpu.Dict type=Pages
<<
        <Count, 2>
        <Kids, [(1 0 R) (8 0 R)]>
        <MediaBox, [0 0 595.27 841.89]>
        <Type, Pages>
>>
    4:   offset=    2828 generation=0 pdfcpu.Dict
<<
        <ColorSpace, <<
                <CS1, DeviceRGB>
        >>>
        <Font, <<
                <G1F18, (6 0 R)>
                <G1F3, (5 0 R)>
                <G1F6, (7 0 R)>
        >>>
        <ProcSet, [PDF Text]>
>>
    5:   offset=    4942 generation=0 pdfcpu.Dict type=Font s
<<
        <BaseFont, Helvetica>
        <Encoding, <<
                <BaseEncoding, WinAnsiEncoding>
                <Differences, [45 minus]>
                <Type, Encoding>
        >>>
        <Name, G1F3>
        <Subtype, Type1>
        <Type, Font>
>>
    6:   offset=    4761 generation=0 pdfcpu.Dict type=Font s
<<
        <BaseFont, Helvetica-Bold>
        <Encoding, <<
                <BaseEncoding, WinAnsiEncoding>
                <Differences, [45 minus]>
                <Type, Encoding>
        >>>
        <Name, G1F18>
        <Subtype, Type1>
        <Type, Font>
>>
    7:   offset=    4578 generation=0 pdfcpu.Dict type=Font s
<<
        <BaseFont, Helvetica-Oblique>
        <Encoding, <<
                <BaseEncoding, WinAnsiEncoding>
                <Differences, [45 minus]>
                <Type, Encoding>
        >>>
        <Name, G1F6>
        <Subtype, Type1>
        <Type, Font>
>>
    8:   offset=    2964 generation=0 pdfcpu.Dict type=Page
```

```
<<
        <Contents, (9 0 R)>
        <Parent, (3 0 R)>
        <Resources, (10 0 R)>
        <Type, Page>
>>
    9:   offset=    3051 generation=0 pdfcpu.StreamDict
<<
        <Filter, LZWDecode>
        <Length, 1316>
>>
   10:   offset=    4441 generation=0 pdfcpu.Dict
<<
        <ColorSpace, <<
                <CS1, DeviceRGB>
        >>>
        <Font, <<
                <G1F18, (6 0 R)>
                <G1F3, (5 0 R)>
                <G1F6, (7 0 R)>
        >>>
        <ProcSet, [PDF Text]>
>>
   11:   offset=    5218 generation=0 pdfcpu.Dict type=Catalo
<<
        <Pages, (3 0 R)>
        <Type, Catalog>
>>
   12:   offset=    5272 generation=0 pdfcpu.Dict
<<
        <Author, ()>
        <CreationDate, (D:20150122062117)>
        <Creator, ()>
        <Keywords, ()>
        <Producer, ()>
        <Subject, ()>
        <Title, (Test)>
>>

Empty free list.

Total pages: 2
```

```
Fonts for page 1:
obj     prefix     Fontname                       Subtype
#5                 Helvetica                      Type1
#6                 Helvetica-Bold                 Type1
#7                 Helvetica-Oblique              Type1

Fonts for page 2:
obj     prefix     Fontname                       Subtype
#5                 Helvetica                      Type1
#6                 Helvetica-Bold                 Type1
#7                 Helvetica-Oblique              Type1

Fontobjects:
obj     prefix     Fontname                       Subtype
#5                 Helvetica                      Type1
#6                 Helvetica-Bold                 Type1
#7                 Helvetica-Oblique              Type1

Fonts:
obj     prefix     Fontname                       Subtype
#5                 Helvetica                      Type1
#6                 Helvetica-Bold                 Type1
#7                 Helvetica-Oblique              Type1

Duplicate Fonts:


No image info available.


writing test_opt.pdf ...
 INFO: 2019/02/20 23:20:12 writing to a.pdf
STATS: 2019/02/20 23:20:12 0 original empty xref entries:
STATS: 2019/02/20 23:20:12 0 original redundant font entries:
STATS: 2019/02/20 23:20:12 0 original redundant image entries
STATS: 2019/02/20 23:20:12 0 original redundant info entries:
STATS: 2019/02/20 23:20:12 0 original objectStream entries:
STATS: 2019/02/20 23:20:12 0 original xrefStream entries:
STATS: 2019/02/20 23:20:12 0 original linearization entries:
STATS: 2019/02/20 23:20:12 XRefTable:
*************************************************************
```

```
HeaderVersion: 1.2
has 2 pages
XRefTable:
                     Size: 15
              Root object: (11 0 R)
              Info object: (12 0 R)
                ID object: [<81C4A57DF6A1E411BD62885083B053CD
XRefTable with 15 entres:
    0: f   next=       0 generation=65535
    1: c => obj:13[0] generation=0
<<
        <Contents, (2 0 R)>
        <Parent, (3 0 R)>
        <Resources, (4 0 R)>
        <Type, Page>
>>
    2:   offset=     102 generation=0 pdfcpu.StreamDict
<<
        <Filter, LZWDecode>
        <Length, 2652>
>>
    3: c => obj:13[7] generation=0
<<
        <Count, 2>
        <Kids, [(1 0 R) (8 0 R)]>
        <MediaBox, [0 0 595.27 841.89]>
        <Type, Pages>
>>
    4: c => obj:13[1] generation=0
<<
        <ColorSpace, <<
                <CS1, DeviceRGB>
        >>>
        <Font, <<
                <G1F18, (6 0 R)>
                <G1F3, (5 0 R)>
                <G1F6, (7 0 R)>
        >>>
        <ProcSet, [PDF Text]>
>>
    5: c => obj:13[2] generation=0
<<
```

```
        <BaseFont, Helvetica>
        <Encoding, <<
                <BaseEncoding, WinAnsiEncoding>
                <Differences, [45 minus]>
                <Type, Encoding>
        >>>
        <Name, G1F3>
        <Subtype, Type1>
        <Type, Font>
>>
    6: c => obj:13[4] generation=0
<<
        <BaseFont, Helvetica-Bold>
        <Encoding, <<
                <BaseEncoding, WinAnsiEncoding>
                <Differences, [45 minus]>
                <Type, Encoding>
        >>>
        <Name, G1F18>
        <Subtype, Type1>
        <Type, Font>
>>
    7: c => obj:13[3] generation=0
<<
        <BaseFont, Helvetica-Oblique>
        <Encoding, <<
                <BaseEncoding, WinAnsiEncoding>
                <Differences, [45 minus]>
                <Type, Encoding>
        >>>
        <Name, G1F6>
        <Subtype, Type1>
        <Type, Font>
>>
    8: c => obj:13[5] generation=0
<<
        <Contents, (9 0 R)>
        <Parent, (3 0 R)>
        <Resources, (10 0 R)>
        <Type, Page>
>>
    9:   offset=    3051 generation=0 pdfcpu.StreamDict
```

```
<<
        <Filter, LZWDecode>
        <Length, 1316>
>>
   10: c => obj:13[6] generation=0
<<
        <ColorSpace, <<
                <CS1, DeviceRGB>
        >>>
        <Font, <<
                <G1F18, (6 0 R)>
                <G1F3, (5 0 R)>
                <G1F6, (7 0 R)>
        >>>
        <ProcSet, [PDF Text]>
>>
   11:   offset=    5218 generation=0 pdfcpu.Dict type=Catalo
<<
        <Pages, (3 0 R)>
        <Type, Catalog>
>>
   12:   offset=    5272 generation=0 pdfcpu.Dict
<<
        <Author, ()>
        <CreationDate, (D:20190220232012+01'00')>
        <Creator, ()>
        <Keywords, ()>
        <ModDate, (D:20190220232012+01'00')>
        <Producer, (pdfcpu v0.1.21)>
        <Subject, ()>
        <Title, ()>
>>
   13:   offset=nil generation=0 pdfcpu.ObjectStreamDict
<<
        <Filter, FlateDecode>
        <First, 45>
        <Length, 327>
        <N, 8>
        <Type, ObjStm>
>>
object stream count:8 size of objectarray:0
   14:   offset=nil generation=0 pdfcpu.XRefStreamDict
```

```
<<
        <Filter, FlateDecode>
        <ID, [<81C4A57DF6A1E411BD62885083B053CD> <e4fcab0bb58
        <Index, [0 14]>
        <Info, (12 0 R)>
        <Length, 63>
        <Root, (11 0 R)>
        <Size, 15>
        <Type, XRef>
        <W, [1 2 2]>
>>

Empty free list.

Total pages: 2

Fonts for page 1:
obj     prefix     Fontname                       Subtype
#5                 Helvetica                      Type1
#6                 Helvetica-Bold                 Type1
#7                 Helvetica-Oblique              Type1

Fonts for page 2:
obj     prefix     Fontname                       Subtype
#5                 Helvetica                      Type1
#6                 Helvetica-Bold                 Type1
#7                 Helvetica-Oblique              Type1

Fontobjects:
obj     prefix     Fontname                       Subtype
#5                 Helvetica                      Type1
#6                 Helvetica-Bold                 Type1
#7                 Helvetica-Oblique              Type1

Fonts:
obj     prefix     Fontname                       Subtype
#5                 Helvetica                      Type1
#6                 Helvetica-Bold                 Type1
#7                 Helvetica-Oblique              Type1

Duplicate Fonts:
```

```
No image info available.


STATS: 2019/02/20 23:20:12 Timing:
STATS: 2019/02/20 23:20:12 read                 :  0.001s  28
STATS: 2019/02/20 23:20:12 validate             :  0.000s   4
STATS: 2019/02/20 23:20:12 optimize             :  0.000s   1
STATS: 2019/02/20 23:20:12 write                :  0.002s  48
STATS: 2019/02/20 23:20:12 total processing time:  0.003s

STATS: 2019/02/20 23:20:12 Original:
STATS: 2019/02/20 23:20:12 File Size            : 6 KB (5884
STATS: 2019/02/20 23:20:12 Total Binary Data    : 4 KB (3968
STATS: 2019/02/20 23:20:12 Total Text   Data    : 2 KB (1916

STATS: 2019/02/20 23:20:12 Breakup of binary data:
STATS: 2019/02/20 23:20:12 images               : 0.000000 By
STATS: 2019/02/20 23:20:12 fonts                : 0.000000 By
STATS: 2019/02/20 23:20:12 other                : 4 KB (3968

STATS: 2019/02/20 23:20:12 Optimized:
STATS: 2019/02/20 23:20:12 File Size            : 5 KB (5034
STATS: 2019/02/20 23:20:12 Total Binary Data    : 4 KB (4358
STATS: 2019/02/20 23:20:12 Total Text   Data    : 676.000000

STATS: 2019/02/20 23:20:12 Breakup of binary data:
STATS: 2019/02/20 23:20:12 images               : 0.000000 By
STATS: 2019/02/20 23:20:12 fonts                : 0.000000 By
STATS: 2019/02/20 23:20:12 other                : 4 KB (4358
```