

PDF tools

An introduction

Martin Schröder

EuroBach $\text{T}_{\text{E}}\text{X}$ 2011

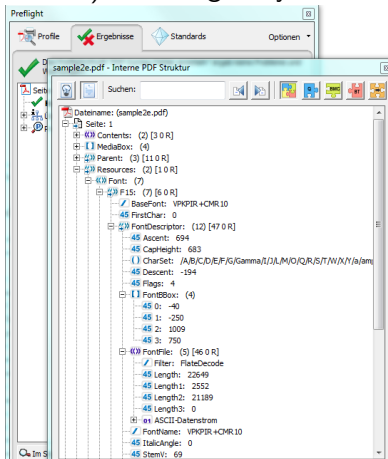
29th April – 3rd May 2011, Bachotek, Poland

Documentation

- ▶ The PDF standard and extensions by Adobe
- ▶ A very good technical introduction to PDF
- ▶ The PDF/A Competence Center
- ▶ Wikipedia
- ▶ OMG WTF PDF – security of PDFs

Adobe Acrobat

Adobe Acrobat is the primary tool for working with PDFs (only on Windows and OSX). It can edit PDFs to a certain degree and allows you to test PDFs (through it's preflight feature). It also gives you access to the structure of a PDF:



PDF tools

Martin Schröder

Documentation

Adobe Acrobat

Plug-Ins

PDF CanOpener

Enfocus PitStop

XPDF/Poppler

QPDF

pdftk

PoDoFoBrowser

diffpdf

pdfsizeopt

Virgo Viewer

OneVision Solvero

PDF CanOpener

PDF CanOpener is an Adobe Acrobat Plug-In for detailed analysis and manipulation of PDF documents aimed at developers. It provides instant access to information about drawn objects (fonts, color spaces, page location, etc.) as well as an object level tree view for analysis and manipulation of the internal object structure.

The screenshot displays a hierarchical tree view of PDF objects. Red arrows point to specific entries with explanatory text:

- Start of 5th Annotation in the Page Annotation List, it's really just a frame for the real movie annotation (points to the first 'AN' object).
- Action Dictionary. Has the Movie Action (points to the 'A' object).
- The real movie Annotation is buried in the Actions here (points to the 'AN' object nested under the 'A' object).
- Another Action Dictionary (points to the 'A' object nested under the second 'AN' object).
- Rendition Object (points to the 'R' object nested under the second 'AN' object).
- (Media Clip Object) (points to the 'M' object nested under the second 'AN' object).
- Embedded FileSpec Dictionary (points to the 'D' object nested under the second 'AN' object).
- Actual QuickTime Movie Data (points to the 'F' object nested under the 'D' object).
- (This shows the Action is a Rendition Action) (points to the 'AN' object nested under the second 'AN' object).

Enfocus PitStop

Enfocus PitStop Pro is an Adobe Acrobat Plug-In for the analysis and manipulation of PDF documents. An unique feature is it's eye dropper that allows the analysis of objects by visually selecting them.

PDF tools

Martin Schröder

Documentation

Adobe Acrobat

Plug-Ins

PDF CanOpener

Enfocus PitStop

XPDF/Poppler

QPDF

pdftk

PoDoFoBrowser

diffpdf

pdfsizeopt

Virgo Viewer

OneVision Solvero

XPDF/Poppler

XPDF (and it's improved version Poppler) includes a number of tools for working with PDFs. These include:

[pdftinfo](#) for basic information about a PDF

[pdffonts](#) gives information about the fonts used by a PDF

[pdftotext](#) converts a PDF to text

[pdftohtml](#) converts a PDF to HTML

[pdfimages](#) extracts images from a PDF

QPDF

QPDF is a command-line program that does structural, content-preserving transformations on PDF files. It can *linearize*, encrypt or decrypt, and check PDFs. It also allows the conversion of PDFs to and from an easily editable format.

PDF tools

Martin Schröder

Documentation

Adobe Acrobat

Plug-Ins

PDF CanOpener

Enfocus PitStop

XPDF/Poppler

QPDF

pdftk

PoDoFoBrowser

diffpdf

pdfsizeopt

Virgo Viewer

OneVision Solvero

pdftk

pdftk is a command-line utility that can merge, split, rotate, unpack (into an editable format) and pack, and repair PDFs (and some other things).

PoDoFoBrowser is a graphical browser and editor for the object structure of PDFs.

Object	Type	Value
/root	Dictionary	<< /Type /Catalog ... >>
1 (x) PTEX.Fullbanner	String	(This is pdfTeX, Version 3.14159-1.10b)
2 (x) Pages	Reference	16 0 R
3 (x) Count	Dictionary	<< /Type /Pages ... >>
4 (x) Kids	Number	1
5 (x) <element 0>	Array	2 0 R
6 (x) 2 0 obj	Reference	<< /Type /Page ... >>
7 (x) Contents	Dictionary	<< /Type /Page ... >>
8 (x) Contents	Dictionary	<< /Type /Page ... >>
9 (x) MediaBox	Array	<< /Type /Page ... >>
10 (x) Parent	Reference	16 0 R

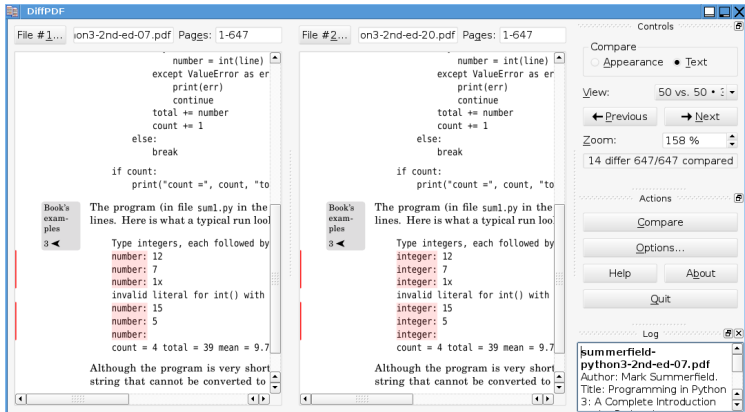
Contents of the associated stream:
Stream associated with 3 0 obj

```
BT
/F17 17.215 Tf 86.4 721.073 Td[(Mee)-1(t)(ng)-374(am)-375(1)15.1(10)1(2)1(00)1(6)]TjF17 14.346 Tf 0 -29.772 Td[(T)93(e)1
(Inehmer)TjF20 11.955 Tf 17.434 -22.167 Td[(v)17]TjF15 11.955 Tf 11.831 0 Td[(Bel-1(nedk)1(t)-327(Hal)1(uptmann)-326
(0)SOfcharef)27(h)1(e)-1(k)1(n)1(051)TjF20 11.955 Tf -11.831 -24.408 Td[(v)17]TjF15 11.955 Tf 11.831 0 Td[(Dal)1(nie)-1(l)-326
(Neub)-27(er)-327(050)1(P)-1(r)1(o)-54(j)-1(ekt)1(l)1(e)-1(r)1(1)1(051)TjF20 11.955 Tf -11.831 -24.409 Td[(v)17]TjF15 11.955 Tf
11.831 0 Td[(Do)1(m)1(n)1(k)-326(Se)-1(ic)27(h)28(ter)-327(050)1(C)-1(h)efen)27(t)28(w)-1(ic)27(k)1(l)1(e)-1(r)1(1)1(051)TjF17 14.346 Tf
-29.265 -34.661 Td[(Ge)-1(m)1(e)1(n)sam)-374(ausg)1(e)-1(f)Tj 131.345 -0.718 Td( J)Tj -0.448 0.718 Td[(uhr)-1(t)1(e)-375(Arb)-32(e)1
(ter)TjF20 11.955 Tf -113.463 -22.167 Td[(v)17]TjF15 11.955 Tf 11.831 0 Td[(E)1(n)1(e)1-327(P)1(rol)-54(j)1(ekt)-326(Z)1(us)-1(a)1(m)1(m)1(n)1(f)-1(a)1
(ss)-1(ung)1-326(f)Tj 169.927 -0.471 Td( J)Tj -0.325 0.471 Td[(ur)TjF21 11.955 Tf 14.958 0 Td[(Co)1(MA)TjF15 11.955 Tf 42.45 0
Td[(wurde)-327(er)1(e)1st)1TjF20 11.955 Tf -238.841 -24.409 Td[(v)17]TjF15 11.955 Tf 11.831 0 Td[(E)1(n)-327(Hal)1(indout)-326(f)Tj
73.969 -0.471 Td( J)Tj -0.325 0.471 Td[(ur)-326(M)1(t)27(w)27(o)1-26(c)27(h)-327(wurde)-326(v)27(or)1(b)-28(er)1(e)-1(r)1(1)1(e)-1

```

PoDoFo also comes with a number of command-line tools for PDFs, among them one for cropping all pages of a PDF. There's also an imposition tool which allows imposition plans written in Lua.

diffpdf can compare two PDFs and display the difference visually.



picture © Qtrac Ltd.

pdfsizeopt

pdfsizeopt is a collection of best practices and scripts for Unix to optimize the size of PDF (≤ 1.4) files.

PDF tools

Martin Schröder

Documentation

Adobe Acrobat

Plug-Ins

PDF CanOpener

Enfocus PitStop

XPDF/Poppler

QPDF

pdftk

PoDoFoBrowser

diffpdf

pdfsizeopt

Virgo Viewer

OneVision Solvero

OneVision Solvero

OneVision Solvero is an interactive PDF, EPS, PS, TIFF and XPS editor for high-end prepress use.

Documentation

Adobe Acrobat

Plug-Ins

PDF CanOpener

Enfocus PitStop

XPDF/Poppler

QPDF

pdftk

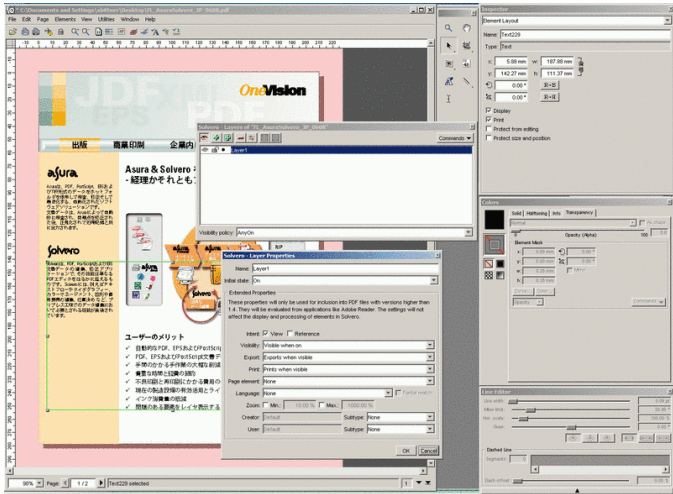
PoDoFoBrowser

diffpdf

pdfsizeopt

Virgo Viewer

OneVision Solvero



picture © OneVision Software AG