NAME

pdfimages – Portable Document Format (PDF) image extractor (version 3.03)

SYNOPSIS

pdfimages [options] *PDF-file image-root*

DESCRIPTION

Pdfimages saves images from a Portable Document Format (PDF) file as Portable Pixmap (PPM), Portable Bitmap (PBM), Portable Network Graphics (PNG), Tagged Image File Format (TIFF), JPEG, JPEG2000, or JBIG2 files.

Pdfimages reads the PDF file *PDF-file*, scans one or more pages, and writes one file for each image, *image-root-nnn.xxx*, where *nnn* is the image number and *xxx* is the image type (.ppm, .pbm, .png, .tif, .jpg, jp2, jb2e, or jb2g).

The default output format is PBM (for monochrome images) or PPM for non-monochrome. The −png or −tiff options change to default output to PNG or TIFF respectively. If both −png and −tiff are specified, CMYK images will be written as TIFF and all other images will be written as PNG. In addition the −j, −jp2, and −jbig2 options will cause JPEG, JPEG2000, and JBIG2, respectively, images in the PDF file to be written in their native format.

OPTIONS

−f *number*
: Specifies the first page to scan.

−l *number*
: Specifies the last page to scan.

−png Change the default output format to PNG.

−tiff Change the default output format to TIFF.

−j Write images in JPEG format as JPEG files instead of the default format. The JPEG file is identical to the JPEG data stored in the PDF.

−jp2 Write images in JPEG2000 format as JP2 files instead of the default format. The JP2 file is identical to the JPEG2000 data stored in the PDF.

−jbig2 Write images in JBIG2 format as JBIG2 files instead of the default format. JBIG2 data in PDF is of the embedded type. The embedded type of JBIG2 has an optional separate file containing global data. The embedded data is written with the extension .jb2e and the global data (if available) will be written to the same image number with the extension .jb2g. The content of both these files is identical to the JBIG2 data in the PDF.

−ccitt Write images in CCITT format as CCITT files instead of the default format. The CCITT file is identical to the CCITT data stored in the PDF. PDF files contain additional parameters specifying how to decode the CCITT data. These parameters are translated to fax2tiff input options and written to a .params file with the same image number. The parameters are:

- **−1** 1D Group 3 encoding
- **−2** 2D Group 3 encoding
- **−4** Group 4 encoding
- **−A** Beginning of line is aligned on a byte boundary
- **−P** Beginning of line is not aligned on a byte boundary
- **−X n** The image width in pixels
- **−W** Encoding uses 1 for black and 0 for white
- **−B** Encoding uses 0 for black and 1 for white
- **−M** Input data fills from most significant bit to least significant bit.

−all Write JPEG, JPEG2000, JBIG2, and CCITT images in their native format. CMYK files are written as TIFF files. All other images are written as PNG files. This is equivalent to specifying the options −png −tiff −j −jp2 −jbig2 −ccitt.

−list Instead of writing the images, list the images along with various information for each image. Do not specify an *image-root* with this option.

The following information is listed for each image:

page the page number containing the image

num the image number

type the image type:

image - an opaque image
mask - a monochrome mask image
smask - a soft-mask image
stencil - a monochrome mask image used for painting a color or pattern

Note: Tranparency in images is represented in PDF using a separate image for the image and the mask/smask. The mask/smask used as part of a transparent image always immediately follows the image in the image list.

width image width (in pixels)

height image height (in pixels)

Note: the image width/height is the size of the embedded image, not the size the image will be rendered at.

color image color space:

gray - Gray
rgb - RGB
cmyk - CMYK
lab - L*a*b
icc - ICC Based
index - Indexed Color
sep - Separation
devn - DeviceN

comp number of color components

bpc bits per component

enc encoding:

image - raster image (may be Flate or LZW compressed but does not use an image encoding)
jpeg - Joint Photographic Experts Group
jp2 - JPEG2000
jbig2 - Joint Bi-Level Image Experts Group
ccitt - CCITT Group 3 or Group 4 Fax

interp "yes" if the interpolation is to be performed when scaling up the image

object ID
the image dictionary object ID (number and generation)

x−ppi The horizontal resolution of the image (in pixels per inch) when rendered on the pdf page.

y−ppi The vertical resolution of the image (in pixels per inch) when rendered on the pdf page.

size The size of the embedded image in the pdf file. The following suffixes are used: 'B' bytes, 'K' kilobytes, 'M' megabytes, and 'G' gigabytes.

ratio The compression ratio of the embedded image.

−opw *password*
Specify the owner password for the PDF file. Providing this will bypass all security restrictions.

−upw *password*
Specify the user password for the PDF file.

−p Include page numbers in output file names.

−q Don't print any messages or errors.

−v Print copyright and version information.

−h Print usage information. (**−help** and **−−help** are equivalent.)

EXIT CODES

The Xpdf tools use the following exit codes:

0 No error.

1 Error opening a PDF file.

2 Error opening an output file.

3 Error related to PDF permissions.

99 Other error.

AUTHOR

SEE ALSO

pdfdetach(1), **pdffonts**(1), **pdfinfo**(1), **pdftocairo**(1), **pdftohtml**(1), **pdftoppm**(1), **pdftops**(1), **pdftotext**(1) **pdfseparate**(1), **pdfsig**(1), **pdfunite**(1)