

**NAME**

pdftohtml – program to convert PDF files into HTML, XML and PNG images

**SYNOPSIS**

**pdftohtml** [*options*] <PDF-file> [<HTML-file> <XML-file>]

**DESCRIPTION**

This manual page documents briefly the **pdftohtml** command. This manual page was written for the Debian GNU/Linux distribution because the original program does not have a manual page.

**pdftohtml** is a program that converts PDF documents into HTML. It generates its output in the current working directory. If *PDF-file* is '-', it reads the PDF file from stdin.

**OPTIONS**

A summary of options are included below.

**-h, -help**

Show summary of options.

**-f <int>**

first page to print

**-l <int>**

last page to print

**-q** do not print any messages or errors

**-v** print copyright and version info

**-p** exchange .pdf links with .html

**-c** generate complex output

**-s** generate single HTML that includes all pages

**-dataurls**

use data URLs instead of external images in HTML. No available in all platforms

**-i** ignore images

**-noframes**

generate no frames. Not supported in complex output mode.

**-stdout**

use standard output

**-zoom <fp>**

zoom the PDF document (default 1.5) (1 means 72 DPI)

**-xml** output for XML post-processing

**-noroundcoord**

do not round coordinates (with XML output only)

**-enc <string>**

output text encoding name

**-opw <string>**

owner password (for encrypted files)

**-upw <string>**

user password (for encrypted files)

**-hidden**

force hidden text extraction

**-fmt** image file format for Splash output (png or jpg). If complex is selected, but **-fmt** is not specified, **-fmt** png will be assumed

**-nomerge**

do not merge paragraphs

**-nodrm**

override document DRM settings

**-wbt <fp>**

adjust the word break threshold percent. Default is 10. Word break occurs when distance between two adjacent characters is greater than this percent of character height.

**-fontfullname**

outputs the font name without any substitutions.

**AUTHOR**

Pdftohtml was developed by Gueorgui Ovtcharov and Rainer Dorsch. It is based and benefits a lot from Derek Noonburg's xpdf package.

This manual page was written by Søren Boll Overgaard <boll@debian.org>, for the Debian GNU/Linux system (but may be used by others).

**SEE ALSO**

**pdfdetach(1), pdffonts(1), pdfimages(1), pdfinfo(1), pdftocairo(1), pdftoppm(1), pdftops(1), pdfto-text(1) pdfseparate(1), pdfsig(1), pdfunite(1)**