NAME

pdftosrc − extract source file or stream from PDF file

SYNOPSIS

pdftosrc *PDF-file* [*stream-object-number*]

DESCRIPTION

If only *PDF-file* is given as argument, **pdftosrc** extracts the embedded source file from the first found stream object with /Type /SourceFile within the *PDF-file* and writes it to a file with the name /SourceName as defined in that PDF stream object (see application example below).

If both *PDF-file* and *stream-object-number* are given as arguments, and *stream-object-number* is positive, **pdftosrc** extracts and uncompresses the PDF stream of the object given by its *stream-object-number* from the *PDF-file* and writes it to a file named *PDF-file.stream-object-number* with the ending **.pdf** or **.PDF** stripped from the original *PDF-file* name.

A special case is related to XRef object streams that are part of the PDF standard from PDF-1.5 onward: If *stream-object-number* equals -1, then **pdftosrc** decompresses the XRef stream from the PDF file and writes it in human-readable PDF cross-reference table format to a file named *PDF-file***.xref** (these XRef streams can not be extracted just by giving their object number).

In any case an existing file with the output file name will be overwritten.

OPTIONS

None.

FILES

Just the executable **pdftosrc**.

ENVIRONMENT

None.

DIAGNOSTICS

At success the exit code of **pdftosrc** is 0, else 1.

All messages go to stderr. At program invocation, **pdftosrc** issues the current version number of the program **xpdf**, on which **pdftosrc** is based:

pdftosrc version 3.01

When **pdftosrc** was successful with the output file writing, one of the following messages will be issued:

Source file extracted to *source-file-name*

or

Stream object extracted to *PDF-file.stream-object-number*

or

Cross-reference table extracted to *PDF-file*.xref

When the object given by the *stream-object-number* does not contain a stream, **pdftosrc** issues the following error message:

      Not a Stream object

When the *PDF-file* can't be opened, the error message is:

      Error: Couldn't open file '*PDF-file*'.

When **pdftosrc** encounters an invalid PDF file, the error message (several lines) is:

      Error: May not be a PDF file (continuing anyway)
      (more lines)
      Invalid PDF file

There are also more error messages from **pdftosrc** for various kinds of broken PDF files.

NOTES

An embedded source file will be written out unchanged, i. e. it will not be uncompressed in this process.

Only the stream of the object will be written, i. e. not the dictionary of that object.

Knowing which *stream-object-number* to query requires information about the PDF file that has to be gained elsewhere, e. g. by looking into the PDF file with an editor.

The stream extraction capabilities of **pdftosrc** (e. g. regarding understood PDF versions and filter types) follow the capabilities of the underlying **xpdf** program version.

Currently the generation number of the stream object is not supported. The default value 0 (zero) is taken.

The wording *stream-object-number* has nothing to do with the 'object streams' introduced by the Adobe PDF Reference, 5th edition, version 1.6.

EXAMPLES

When using pdftex, a source file can be embedded into some *PDF-file* by using pdftex primitives, as illustrated by the following example:

```
\immediate\pdfobj
    stream attr {/Type /SourceFile /SourceName (myfile.zip)}
    file{myfile.zip}
\pdfcatalog{/SourceObject \the\pdflastobj\space 0 R}
```

Then this zip file can be extracted from the *PDF-file* by calling **pdftosrc** *PDF-file*.

BUGS

Not all embedded source files will be extracted, only the first found one.

Email bug reports to **pdftex@tug.org.**

SEE ALSO

xpdf(1), **pdfimages**(1), **pdftotext**(1), **pdftex**(1),

AUTHORS

pdftosrc written by Han The Thanh, using **xpdf** functionality from Derek Noonburg.

Man page written by Hartmut Henkel.

COPYRIGHT

Copyright (c) 1996-2006 Han The Thanh, <thanh@pdftex.org>

This file is part of pdfTeX.

pdfTeX 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 2 of the License, or (at your option) any later version.

pdfTeX 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 pdfTeX; if not, write to the Free Software Foundation, Inc., 59 Temple Place, Suite 330, Boston, MA 02111-1307 USA