

PDFGREP(1) Pdftotext Manual PDFGREP(1)

NAME pdftotext - search pdf files for a regular expression

SYNOPSIS pdftotext [OPTION...] PATTERN FILE...

DESCRIPTION Search for PATTERN in each FILE. PATTERN is an extended regular expression.

pdftotext works much like grep, with one distinction: It operates on pages and not on lines.

OPTIONS -i, --ignore-case Ignore case distinctions in both the PATTERN and the input files.

- H, --with-filename
Print the file name for each match. This is the default setting when there is more than one file to search.
- h, --no-filename
Suppress the prefixing of file name on output. This is the default setting when there is only one file to search.
- n, --page-number
Prefix each match with the number of the page where it was found.
- c, --count
Suppress normal output. Instead print the number of matches for each input file. Note that unlike grep, multiple matches on the same page will be counted individually.
- p, --page-count
Like -c, but prints the number of matches per page.
- C, --context NUM
Print at most INUM characters of context around each match. The exact number will vary, because pdftotext tries to respect word boundaries. If NUM is "line", the whole line will be printed. If this option is not set, pdftotext tries to print lines that are not longer than the terminal width.
- color WHEN
Surround file names, page numbers and matched text with escape sequences to display them in color on the terminal. (The default setting is auto). WHEN can be:

always

Always use colors, even when stdout is not a terminal.

never

Do not use colors.

auto

Use colors only when stdout is a terminal.

`-r, --recursive`

Recursively search all files (restricted by `--include` and `--exclude`) under each directory, following symlinks only if they are on the command line.

`-R, --dereference-recursive`

Same as `-r`, but follows all symlinks.

`--exclude=GLOB`

Skip files whose base name matches GLOB. See `glob(7)` for wildcards you can use. You can use this option multiple times to exclude more patterns. It takes precedence over `--include`. Note, that in- and excludes apply only to files found via `--recursive` and not to the argument list.

`--include=GLOB`

Only search files whose base name matches GLOB. See `--exclude` for details. The default is `*.pdf`.

`--password=PASSWORD`

Use PASSWORD to decrypt the PDF-files. Note that this password will show up in your command history and the output of `ps(1)`. So please do not use this if the security of PASSWORD is important. It is also currently not possible to use individual passwords for different files.

`-m, --max-count NUM`

Stop reading a file after NUM matching lines. When the `-c` or `--count` option is also used, `pdftopdf` does not output a count greater than NUM.

`--unac`

Remove accents and ligatures from both the search pattern and the PDF documents. This is useful if you want to search for a word containing "ae", but the PDF uses the single character "æ" instead. See `unac(3)` and `unaccent(1)` for details.

This option is experimental and only available if `pdftopdf` is

compiled with unac support.

`-q, --quiet`

Suppress all normal output to stdout. Errors will be printed and the exit codes will be returned (see below).

`--help`

Print a short summary of the options.

`-V, --version` Show version information.

ENVIRONMENT VARIABLES The behavior of `pdgrep` is affected by the following environment variable.

GREP_COLORS

Specifies the colors and other attributes used to highlight various parts of the output. The syntax and values are like `GREP_COLORS` of `grep`. See `grep(1)` for more details. Currently only the capabilities `mt`, `ms`, `mc`, `fn`, `ln` and `se` are used by `pdgrep`, where `mt`, `ms` and `mc` have the same effect on `pdgrep`.

EXIT STATUS Normally, the exit status is 0 if at least one match is found, 1 if no match is found and 2 if an error occurred. But if the `--quiet` or `-q` option is used and a match was found, `pdgrep` will return 0 regardless of errors.

AUTHOR Hans-Peter Deifel

SEE ALSO `grep(1)`, `regex(7)`

`Pdgrep 1.3.1 08/10/2014 PDFGREP(1)`