

# Package: Rpoppler (via r-universe)

September 19, 2024

**Version** 0.1-3

**Title** PDF Tools Based on Poppler

**Description** PDF tools based on the Poppler PDF rendering library. See <http://poppler.freedesktop.org/> for more information on Poppler.

**License** GPL-2

**SystemRequirements** Poppler Glib interface headers and libraries (<http://poppler.freedesktop.org/>) [Debian/Ubuntu: libpoppler-glib-dev, Fedora: poppler-glib-devel]

**NeedsCompilation** yes

**Author** Kurt Hornik [aut, cre] (<https://orcid.org/0000-0003-4198-9911>)

**Maintainer** Kurt Hornik <Kurt.Hornik@R-project.org>

**Repository** CRAN

**Date/Publication** 2024-08-13 11:13:31 UTC

## Contents

PDF_doc	1
PDF_fonts	2
PDF_info	3
PDF_text	3

<b>Index</b>	<b>5</b>
--------------	----------

---

PDF_doc	<i>PDF document reference</i>
---------	-------------------------------

---

## Description

Create a reference to a Portable Document Format (PDF) file for use in subsequent information extraction from the file.

**Usage**

```
PDF_doc(file)
```

**Arguments**

`file`            A character string giving the path to a PDF file.

**Value**

A reference to a PDF file (external pointer object).

**Examples**

```
file <- system.file(file.path("doc", "Sweave.pdf"), package = "utils")
doc <- PDF_doc(file)
## Can now use the reference for information extraction, avoiding
## the creation of new PopplerDocument objects when doing so.
PDF_info(doc)
PDF_fonts(doc)
```

---

PDF\_fonts

*PDF font information*

---

**Description**

Obtain the fonts used in a Portable Document Format (PDF) file and further information about these fonts.

**Usage**

```
PDF_fonts(file)
```

**Arguments**

`file`            A character string giving the path to a PDF file, or an object of class "PDF\_doc" giving a reference to a PDF file.

**Value**

A data frame inheriting from PDF\_fonts (which has a useful print method), with the following variables:

<code>name</code>	the full name of the font (character)
<code>type</code>	the font type (Type 1, Type 3, etc.; character)
<code>file</code>	the file name of the font (character; empty if the font is embedded)
<code>emb</code>	whether the font is embedded in the PDF file or not (logical)
<code>sub</code>	whether the font is a subset of another font (logical)

**Examples**

```
file <- system.file(file.path("doc", "Sweave.pdf"), package = "utils")
PDF_fonts(file)
```

---

PDF_info	<i>PDF document information</i>
----------	---------------------------------

---

**Description**

Extract document information from a Portable Document Format (PDF) file.

**Usage**

```
PDF_info(file)
```

**Arguments**

**file** A character string giving the path to a PDF file, or an object of class "PDF\_doc" giving a reference to a PDF file.

**Value**

An object of class PDF\_info (which has useful format and print methods), containing the information in the PDF Info dictionary (title, subject, keywords, author, creator, producer, creation date, modification date) as well as the number of pages and the page sizes, whether the document is optimized (linearized), and the PDF version it uses.

**Examples**

```
file <- system.file(file.path("doc", "Sweave.pdf"), package = "utils")
PDF_info(file)
```

---

PDF_text	<i>PDF text extraction</i>
----------	----------------------------

---

**Description**

Extract text from a Portable Document Format (PDF) file.

**Usage**

```
PDF_text(file)
```

**Arguments**

**file** A character string giving the path to a PDF file, or an object of class "PDF\_doc" giving a reference to a PDF file.

**Value**

A character vector with the extracted texts for each page.

**Examples**

```
file <- system.file(file.path("doc", "Sweave.pdf"), package = "utils")
PDF_text(file)
```

# Index

PDF\_doc, [1](#), [2](#), [3](#)

PDF\_fonts, [2](#)

PDF\_info, [3](#)

PDF\_text, [3](#)