

Package ‘Rpoppler’

July 2, 2014

Version 0.0-1

Title PDF tools based on Poppler

Description PDF tools based on the Poppler library

Author Kurt Hornik

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

License GPL-2

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

OS_type unix

NeedsCompilation yes

Repository CRAN

Date/Publication 2013-08-20 15:41:21

R topics documented:

PDF_doc	2
PDF_fonts	2
PDF_info	3
PDF_text	4

Index	5
--------------	----------

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	<i>PDF font information</i>
-----------	-----------------------------

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)
```

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

Description

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

Usage

```
PDF_info(file)
```

Arguments

<code>file</code>	A character string giving the path to a PDF file, or an object of class <code>"PDF_doc"</code> 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

PDF text extraction

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, [2](#), [2](#), [3](#), [4](#)

PDF_fonts, [2](#)

PDF_info, [3](#)

PDF_text, [4](#)