

Package: rmdwc (via r-universe)

September 5, 2024

Type Package

Title Count Words, Chars and Non-Whitespace Chars in R Markdown Docs

Version 0.3.0

Date 2022-11-12

Description If you are using R Markdown documents then you have sometimes restrictions about the size of the documents, e.g. number of words, number of characters or non-whitespace characters. `rmdcount()` computes these counts with and without code chunks and returns the result as data frame.

Imports knitr, rstudioapi

License GPL-3

Encoding UTF-8

RoxygenNote 7.2.1

Roxygen list(markdown = TRUE)

Suggests testthat

NeedsCompilation no

Author Sigbert Klinke [aut, cre]

Maintainer Sigbert Klinke <sigbert@hu-berlin.de>

Date/Publication 2021-09-07 12:00:01 UTC

Repository <https://sigbertklinke.r-universe.dev>

RemoteUrl <https://github.com/sigbertklinke/rmdwc>

RemoteRef HEAD

RemoteSha 7bd5ddb137d0726ac902414558057e6b85fbd121

Contents

<code>rmdcount</code>	2
<code>rmdcountAddin</code>	4
<code>rmdwcl</code>	4

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

rmdcount

*Word, character and non-whitespace characters count***Description**

rmdcount counts lines, words, bytes, characters and non-whitespace characters in R Markdown files excluding code chunks. txtcount counts lines, words, bytes, characters and non-whitespace characters in plain text files.

Note that the counts may differ a bit from unix wc and Libre Office because it depends on the definition of a line, a word and a character.

Usage

```
rmdcount(
  files = NULL,
  space = "[[:space:]]",
  word = "[[:space:]]+",
  line = "\n",
  exclude = "`\`\\\{.*?\`"
)
```

```
txtcount(
  files = NULL,
  space = "[[:space:]]",
  word = "[[:space:]]+",
  line = "\n"
)
```

Arguments

files	character: file name(s)
space	character: pattern to split a text at spaces (default: '[[:space:]]')
word	character: pattern to split a text at word boundaries (default: '[[:space:]]+')
line	character: pattern to split lines (default: '\n')
exclude	character: pattern to exclude text parts, e.g. code chunks (default: '`\`\\\{.*?\`')

Details

We define:

Line the number of lines. It differs from unix wc -l since wc counts the number of newlines.

Word it is considered to be a character or characters delimited by white space. However, a "word" is in general a fuzzy concept, for example is "3.141593" a word? Therefore different programs may count differently, for more details see the discussion to the Libreoffice bug [Word count gives wrong results - Another Example](#) Comment 5.

The following approach is used to detect lines, words, characters and non-whitespace characters.

lines `strsplit(rmd, line)[[1]]` with `line='\n'`

bytes `charToRaw(rmd)`

words `strsplit(rmd, word)[[1]]` with `word='[:space:]+'`

characters `strsplit(rmd, '')[[1]]`

non-whitespace characters `strsplit(gsub(space, '', rmd), '')[[1]]` with `space='[:space:]'`

If `txtcount` is used then code chunks are deleted with `gsub('```\{\.*?\}```', '', rmd)` before counting.

Value

a data frame with following elements

file basename of file

lines number of lines

words number of words

bytes number of bytes

chars number of characters

nonws number of non-whitespace characters

path path of file

Examples

```
# count excluding code chunks
files <- system.file('rmarkdown/rstudio_pdf.Rmd', package="rmdwc")
rmdcount(files)
# count including code chunks
txtcount(files) # or rmdcount(files, exclude='')
# count for a set of R Markdown docs
files <- list.files(path=system.file('rmarkdown', package="rmdwc"),
                   pattern="*.Rmd", full.names=TRUE)
rmdcount(files)
# use of rmdcount() in a R Markdown document
if (interactive()) {
  files <- system.file('rmarkdown/rstudio_pdf.Rmd', package="rmdwc")
  file.edit(files) # SAVE(!) the file and knit it
}
# count including code chunks
files <- system.file('rmarkdown/rstudio_pdf.Rmd', package="rmdwc")
txtcount(files)
```

rmdcountAddin	<i>rmdcountAddin</i>
---------------	----------------------

Description

Applies rmdcount to the current R Markdown document

Usage

```
rmdcountAddin()
```

Value

nothing

Examples

```
if (interactive()) rmdcountAddin()
```

rmdwcl	<i>Word-, character and non-whitespace characters count for a text</i>
--------	--

Description

Counts words, characters and non-whitespace characters in a string. Is used in rmdcount, see details there.

Usage

```
rmdwcl(rmd, space = "[[:space:]]", word = "[[:space:]]+", line = "\n")
```

Arguments

rmd	character: R Markdown document as string
space	character: pattern to split a text at spaces (default: '[[:space:]]')
word	character: pattern to split a text at word boundaries (default: '[[:space:]]+')
line	character: pattern to split lines (default: '\n')

Value

a list

Examples

```
file <- system.file('rmarkdown/rstudio_pdf.Rmd', package="rmdwc")
fcont <- readChar(file, file.info(file)$size)
rmdwcl(fcont)
```

Index

rmcount, [2](#)

rmcountAddin, [4](#)

rmwcl, [4](#)

txtcount (rmcount), [2](#)