

Package ‘ralger’

December 22, 2019

Type Package

Title Easy Web Scraping

Version 1.0.0

Author Mohamed El Fodil Ihaddaden

Maintainer Mohamed El Fodil Ihaddaden <ihaddaden.fodeil@gmail.com>

Description The goal of 'ralger' is to facilitate web scraping in R.
The user has the ability to extract a vector with `scrap()` or a tidy dataframe using `tidy_scrap()`.

License MIT + file LICENSE

Encoding UTF-8

LazyData true

URL <https://github.com/feddelegrand7/ralger>

BugReports <https://github.com/feddelegrand7/ralger/issues>

VignetteBuilder knitr

Imports magrittr, rvest, xml2, testthat, tidyr, dplyr, stringr

Suggests knitr

RoxygenNote 6.1.1

NeedsCompilation no

Repository CRAN

Date/Publication 2019-12-22 16:30:02 UTC

R topics documented:

scrap	2
tidy_scrap	2
Index	4

scrap

Simple website scraping

Description

This function is used to scrap one element from a website.

Usage

```
scrap(link, node, clean = FALSE)
```

Arguments

link	the link of the webpage to scrap
node	the CSS element to consider, the SelectorGadget tool is highly recommended
clean	logical. Should the function clean the extracted vector or not ? Default is FALSE.

Value

a character vector

Examples

```
# Extracting imdb top 250 movie titles  
  
link <- "https://www.imdb.com/chart/top/"  
node <- ".titleColumn a"  
  
scrap(link, node)
```

tidy_scrap*Website Tidy scraping*

Description

This function is used to scrap a tibble from a website.

Usage

```
tidy_scrap(link, nodes, colnames, clean = FALSE)
```

Arguments

<code>link</code>	the link of the webpage to scrap
<code>nodes</code>	the vector of CSS elements to consider, the SelectorGadget tool is highly recommended
<code>colnames</code>	the names of the expected columns
<code>clean</code>	logical. Should the function clean the extracted tibble or not? Default is FALSE.

Value

a tidy dataframe

Examples

```
# Extracting imdb movie titles and rating

link      <- "https://www.imdb.com/chart/top/"
my_nodes <- c(".titleColumn a", "strong")
names     <- c("title", "rating")

tidy_scrap(link, my_nodes, names)
```

Index

scrap, [2](#)

tidy_scrap, [2](#)