

Package ‘scrappy’

May 9, 2026

Title A Simple Web Scraper

Version 0.0.1

Description A group of functions to scrape data from different websites, for academic purposes.

License MIT + file LICENSE

Encoding UTF-8

LazyData true

RoxygenNote 7.1.1

URL <https://github.com/villegar/scrappy/>,
<https://villegar.github.io/scrappy/>

BugReports <https://github.com/villegar/scrappy/issues/>

Language en-GB

Imports magrittr, rvest, xml2

NeedsCompilation no

Author Roberto Villegas-Diaz [aut, cre] (ORCID:
<<https://orcid.org/0000-0001-5036-8661>>)

Maintainer Roberto Villegas-Diaz <villegas.roberto@hotmail.com>

Repository CRAN

Date/Publication 2021-01-09 14:20:02 UTC

Contents

newa_nrcc	2
Index	4

 newa_nrcc

 Retrieve data from NEWA at Cornell University

Description

Retrieve Weather data from the Network for Environment and Weather Applications (NEWA) at Cornell University.

Usage

```
newa_nrcc(
  client,
  year,
  month,
  station,
  base = "http://newa.nrcc.cornell.edu/newaLister",
  interval = "hly",
  sleep = 6,
  table_id = "#dtable",
  path = getwd(),
  save_file = TRUE
)
```

Arguments

client	RSelenium client.
year	Numeric value with the year.
month	Numeric value with the month.
station	String with the station abbreviation. Check the http://newa.cornell.edu/index.php?page=station-pages for a list.
base	Base URL (default: http://newa.nrcc.cornell.edu/newaLister).
interval	String with data interval (default: hly, hourly).
sleep	Numeric value with the number of seconds to wait for the page to load the results (default: 6 seconds).
table_id	String with the unique HTML ID assigned to the table containing the data (default: #dtable)
path	String with path to location where CSV files should be stored (default: getwd()).
save_file	Boolean flag to indicate whether or not the output should be stored as a CSV file.

Value

Tibble with the data retrieved from the server.

Examples

```
## Not run:  
# Create RSelenium session  
rD <- RSelenium::rsDriver(browser = "firefox", port = 4544L, verbose = FALSE)  
# Retrieve data for the Geneva (Bejo) station on 2020/12  
scrappy::nawa_nrcc(rD$client, 2020, 12, "gbe")  
# Stop server  
rD$server$stop()  
  
## End(Not run)
```

Index

nwa_nrcc, 2