Package ‘LIHKGr’

September 8, 2020

Title  Scrape Text Data from LIHKG
Version  0.1.0
Description  The goal of ‘LIHKGr’ is to scrape text data from LIHKG (<https://lihkg.com/>), the Hong Kong version of Reddit. LIHKG is currently protected by Google’s 'reCAPTCHA', this package currently builds on 'RSelenium' and adopts a semi-manual approach to bypass it.
License  GPL-3
URL  https://github.com/justinchuntingho/LIHKGr
BugReports  https://github.com/justinchuntingho/LIHKGr/issues
Encoding  UTF-8
LazyData  true
RoxygenNote  7.1.1.9000
Imports  R6, purrr, tibble (>= 3.0.3), dplyr, RSelenium, raster, magrittr, rvest, xml2
NeedsCompilation  no
Author  Justin Chun-ting Ho [aut, cre]
   (<https://orcid.org/0000-0002-7884-1059>)
Maintainer  Justin Chun-ting Ho <Justin.Ho@ed.ac.uk>
Repository  CRAN
Date/Publication  2020-09-08 08:00:03 UTC

R topics documented:

create_lihkg ................................................................. 2

Index  3

1
create_lihkg

Create a lihkg scraper

Description

This function creates a lihkg scraper

Usage

create_lihkg(..., verbose = FALSE)

Arguments

... parameters to be passed to RSelenium::rsDriver(). For example, you can use create_lihkg(browser = "firefox", port = sample(10000:60000, 1)) to generate a Selenium instance of Firefox.

verbose logical, whether debug information is printed out.

Value

a lihkg scraper

Examples

## Not run:
lihkg <- create_lihkg(browser = "firefox", port = sample(10000:60000, 1))

## End(Not run)
Index

create_lihkg, 2