

Package: webbotparseR (via r-universe)

August 21, 2024

Title Parse html files containing search engine results

Version 0.0.1.9000

Description Parse search engine results which have been scraped with the 'WebBot' browser extension
<<https://github.com/gesiscss/WebBot>>.

URL <https://gesistsa.github.io/webbotparseR/>,
<https://github.com/gesistsa/webbotparseR>

BugReports <https://github.com/gesistsa/webbotparseR/issues>

License MIT + file LICENSE

Encoding UTF-8

Roxygen list(markdown = TRUE)

RoxygenNote 7.2.3

Imports rvest, tibble, fastmap, base64enc

Depends R (>= 3.5)

Suggests covr, testthat (>= 3.0.0)

Config/testthat/edition 3

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

RemoteUrl <https://github.com/schochastics/webbotparseR>

RemoteRef HEAD

RemoteSha 96fe77dcb84595c3f227a2eb6e5e2da775d9e692

Contents

base64_to_img	2
parse_metadata	2
parse_search_results	3

Index	4
--------------	----------

base64_to_img *Image data uri to file*

Description

Convert a data uri to an image in the correct format and save it to a file.

Usage

```
base64_to_img(data_uri, slug)
```

Arguments

data_uri character, base64 image string as returned by [parse_search_results](#)
slug character, name of file to export image to. WITHOUT extension

Value

nothing, called for side effects

Examples

```
## Not run:  
data_uri <- paste0("data:image/png;base64,",  
  base64enc::base64encode(system.file('logo.png', package = "webbotparseR")))  
base64_to_img(data_uri,"logo")  
  
## End(Not run)
```

parse_metadata *Parse metadata from search engine results*

Description

Parse metadata from search engine results

Usage

```
parse_metadata(path)
```

Arguments

path character. a path to a file that contains search results

Value

a tibble of parsed search engine results

Examples

```
parse_metadata("www.google.com_climate_change_text_2023-03-16_08_16_11.html")
```

parse_search_results *Parse search engine results*

Description

Parse search engine results

Usage

```
parse_search_results(path, engine, selectors = "latest")
```

Arguments

path	character. either a path to a file that contains search results or a path to a directory containing search engine result files
engine	character.
selectors	either character or a webbot_selectors S3 object. For character, it represents the selectors version and valid choices are listed in selectors_versions and "latest" (select the latest version). You can also supply your own webbot_selectors object.

Value

a tibble of parsed search engine results

Examples

```
search_html <- system.file(  
  "www.google.com_climatechange_text_2023-03-16_08_16_11.html",  
  package = "webbotparseR")  
  
parse_search_results(search_html, engine = "google text")
```

Index

`base64_to_img`, [2](#)

`parse_metadata`, [2](#)

`parse_search_results`, [2](#), [3](#)