
scrapper_helpers Documentation

Release 0.2.0

Limebrains

Nov 15, 2017

Contents

| | | |
|----------|----------------------------|----------|
| 1 | Utils methods | 3 |
| | Python Module Index | 5 |

Contents:

CHAPTER 1

Utils methods

`scraper_helpers.utils.caching (key_func=<function default_key_func>)`

A decorator that creates local dumps of the decorated function's return values for given parameters. It can take a `key_func` argument that determines the name of the output file.

`scraper_helpers.utils.default_key_func (*args)`

This method creates the default string representation of the input parameters, used for caching filename

`scraper_helpers.utils.finder (many=True, *finder_args, **finder_kwargs)`

Search for an element(or elements depending on variable 'many') with certain key(tag, class, id, ...) in a web page markup

Parameters

- **many** – decide whether searching for one or more elements
- **finder_args** – key to find

Returns element or list of elements

`scraper_helpers.utils.flatten (container)`

Flattens a list :param container: list with nested lists :type container: list :return: list with elements that were nested in container :rtype: list

`scraper_helpers.utils.get_number_from_string (s, number_type, default)`

Returns a numeric value of `number_type` created from the given string or default if the cast is not possible.

`scraper_helpers.utils.get_random_user_agent ()`

Randoms user agent to prevent "python" user agent :return: Random user agent from `USER_AGENTS` :rtype: str

`scraper_helpers.utils.html_decode (s)`

Returns the ASCII decoded version of the given HTML string. This does NOT remove normal HTML tags like `<p>`.

`scraper_helpers.utils.key_md5 (*args)`

This method creates an MD5 from the input parameters, used for caching filename

`scraper_helpers.utils.key_sha1(*args)`

This method creates an SHA-1 from the input parameters, used for caching filename

`scraper_helpers.utils.normalize_text(text, lower=True, replace_spaces='_')`

This method returns the input string, but normalizes it for use in the url. :param text: input string :rtype: string
:return: Normalized string. lowercase, no diacritics, '-' instead of ' '

`scraper_helpers.utils.replace_all(text, dic)`

This method returns the input string, but replaces its characters according to the input dictionary. :param text: input string :param dic: dictionary containing the changes. key is the character that's supposed to be changed and value is

the desired value

Return type string

Returns String with the according characters replaced

S

`scraper_helpers.utils`, 3

C

catching() (in module `scraper_helpers.utils`), 3

D

default_key_func() (in module `scraper_helpers.utils`), 3

F

finder() (in module `scraper_helpers.utils`), 3

flatten() (in module `scraper_helpers.utils`), 3

G

get_number_from_string() (in module `scraper_helpers.utils`), 3

get_random_user_agent() (in module `scraper_helpers.utils`), 3

H

html_decode() (in module `scraper_helpers.utils`), 3

K

key_md5() (in module `scraper_helpers.utils`), 3

key_sha1() (in module `scraper_helpers.utils`), 3

N

normalize_text() (in module `scraper_helpers.utils`), 4

R

replace_all() (in module `scraper_helpers.utils`), 4

S

`scraper_helpers.utils` (module), 3