

Package ‘WikiSocio’

February 23, 2016

Title A MediaWiki API Wrapper

Version 0.7.0

Author person(`Leo", `Joubert", email = `joubert.leo@gmail.com", role =
c(`aut", `cre"))

Maintainer Leo Joubert <joubert.leo@gmail.com>

Description MediaWiki is wiki platform. Providing the infrastructure of Wikipedia, it also offers very sophisticated archiving functionalities. This package is built to store these wiki's archives to R object - data-frame, lists, vector and variables. All data are downloaded with the help of MediaWiki REST API. For instance, you can get all revisions made by a contributor - `contrib_list()`, all the revisions of a page `page_revisions()`, or create corpus of contributors - `corpus_contrib_create()` - and pages `corpus_page_create()`. Then, you can make these corpus rich of data about contributors or pages - `corpus_contrib_data()` or `corpus_page_data()`.

Depends R (>= 3.0.0)

License GPL-3

LazyData true

Imports httr,igraph,pbapply,plyr,XML,RCurl,stringr

RoxygenNote 5.0.1

NeedsCompilation no

Repository CRAN

Date/Publication 2016-02-23 19:25:38

R topics documented:

<code>contrib_list</code>	2
<code>corpus_contrib_create</code>	3
<code>corpus_contrib_data</code>	4
<code>corpus_contrib_select</code>	5
<code>corpus_page_create</code>	6
<code>corpus_page_data</code>	7

graph_page	8
page_anon	8
page_cat	9
page_infos	10
page_islink	10
page_links	11
page_ranking	12
page_revisions	12
period	13
xpath	14

Index	15
--------------	-----------

contrib_list	<i>Getting the list of the contribution of a contributor.</i>
--------------	---

Description

Getting the list of the contribution of a contributor.

Usage

```
contrib_list(x, namespace = "0", domain = "fr", page = F)
```

Arguments

x	The name of the contributor
namespace	The namespace of contributions
domain	The domain of the wiki
page	If TRUE, contrib_list return only a list of all different page in witch the contribors made a modification. If FALSE, return a data-frame with all the contributions, their timestamp and their weight

Value

Depending the value of page, weither a character vector containing the names of all contributors, or a data-frame containing all the revisions with the name of the contributor, a timestamp and a weight

Examples

```
# All the contribution of an user of the french wiki.
contrib_list('cafeine05')

# Return a character vector with all the page modified by this contributor.
contrib_list('cafeine05',page=TRUE)
```

corpus_contrib_create *Creating a contributor corpus*

Description

Creating a contributor corpus

Usage

```
corpus_contrib_create(method, x, limit = "max", domain = "fr")
```

Arguments

method	The method employed to get data into the corpus. category Every contributor page included in the category x would be included in the corpus random Select x random contributor page xpath x is a vector containing in first position the URL of one page, and in second position the xpath request to apply to this URL. Will return a list of words, wich are the names of the contributors if the xpath request is writed correctly
x	The pointer of the method.
limit	If not 'max', then an integer giving the corpus size. Useless for random method.
domain	The domain where the wiki is located.

Value

A character vector

See Also

[corpus_contrib_data](#) to put data into the corpus you made with this function, [corpus_contrib_select](#) to clean the corpus with a specific criteria

Other corpus functions: [corpus_contrib_data](#), [corpus_contrib_select](#), [corpus_page_create](#), [corpus_page_data](#)

Examples

```
# Creating a contributor corpus formed by contributors of the 'Action' page on the french wiki.  
corpus_contrib_create('page', 'Action')
```

```
# Same as previous, by limiting the size of the corpus at 5 contributors.  
corpus_contrib_create('page', 'Action', limit=5)
```

corpus_contrib_data *Adding data to a contributor corpus*

Description

Adding data to a contributor corpus

Usage

```
corpus_contrib_data(method, x, prefix = "Utilisateur", domain = "fr")
```

Arguments

method	The method employed to get data into the corpus. raw For each user, downloading characteristics and testing other member of the corpus. page For each user, downloading edited pages.
x	The users' corpus.
prefix	In the wiki, the prefix for users' pages namespace.
domain	The domain where the wiki is located.

Value

A data-frame.

See Also

Other corpus functions: [corpus_contrib_create](#), [corpus_contrib_select](#), [corpus_page_create](#), [corpus_page_data](#)

Examples

```
c <- corpus_contrib_create('page', 'Action', limit=5)
c <- corpus_contrib_data('raw', c)
```

corpus_contrib_select *Selecting the member of a corpus, based on their contributions.*

Description

This function is used to have a corpus of big contributors of the wiki. To use this, you need a contributor corpus completed with 'page' method of corpus_contrib_data

Usage

```
corpus_contrib_select(method, x, threeshold, domain = "fr")
```

Arguments

method	Method employed to get data into the corpus. firstContrib Select only contributors who are part of the threeshold % of first contributor of each page
x	A corpus created with corpus_contrib_create
threeshold	An integer used as a threeshold to decide whether a contributor can be on the corpus or not. For instance, if threeshold=5, only contributor who are in the 5 percent first contributor of the article.
domain	The domain where the wiki is located

Details

BE CAUTIOUS : this function is VERY time-consuming.

Value

A data-frame.

See Also

Other corpus functions: [corpus_contrib_create](#), [corpus_contrib_data](#), [corpus_page_create](#), [corpus_page_data](#)

Examples

```
# creating a corpus of 5 contributor of the action page
c <- corpus_contrib_create('page','Action',limit=5)

c <- corpus_contrib_data('page',c)

# Keeping on this corpus only the contributor who are part of
# the 5 percent first contributors of the correspondant article.
c <- corpus_contrib_select(c,5)
```

corpus_page_create *Creating a page corpus*

Description

Creating a page corpus

Usage

```
corpus_page_create(method, x, limit = "max", domain = "fr")
```

Arguments

method	The method employed to create the corpus. category Every page in the category <i>x</i> is included to the corpus random Select <i>x</i> random page xpath <i>x</i> is a vector containing in first position the URL of one page, and in second position the wpath request to apply to this URL. Will return a list of words, wich are the names of pages if the xpath request is writed correctly
<i>x</i>	The pointer of the method.
limit	If not 'max', then an integer giving the length of the corpus. Useless for random method.
domain	The domain where the wiki is located

Value

A character vector.

See Also

[corpus_contrib_create](#) for contributors corpus, [corpus_page_data](#) to put data into the corpus you made with this function

Other corpus functions: [corpus_contrib_create](#), [corpus_contrib_data](#), [corpus_contrib_select](#), [corpus_page_data](#)

Examples

```
# Creating a page corpus formed by 3 random selected page
corpus_page_create('random',3)
```

corpus_page_data *Adding page data to a page corpus*

Description

Adding page data to a page corpus

Usage

```
corpus_page_data(method, x, selection = c("nbLinks", "nbContrib",
    "nbRevisions", "percentAnon", "percent10"), domain = "fr")
```

Arguments

method	The method employed to get data into the corpus. variables For each page of the corpus, return variables specified in selection
x	A character vector created with corpus_page_create
selection	A character vector giving all the data variables to get in the corpus : nbLinks Number of links in the page nbContrib Number of contributors who have edited the page at least once nbRevisions Number of revisions of the page percentAnon Percentage of anonymous contributions percent10 Percentage of the text due to the 10 % first contributor
domain	The domain where the wiki is located

Value

A data-frame.

See Also

Other corpus functions: [corpus_contrib_create](#), [corpus_contrib_data](#), [corpus_contrib_select](#), [corpus_page_create](#)

Examples

```
# Creating a page corpus with 3 randomly selected page
corpus <- corpus_page_create('random',3)
corpus <- corpus_page_data("variables",corpus)
```

graph_page	<i>Graph a wiki page</i>
------------	--------------------------

Description

Creating a graph objet in witch an edge is a wiki page and a vertice is a link beetween two wiki pages.

Usage

```
graph_page(x, domain = "fr", namespace = "0")
```

Arguments

x	Can be either a string defining a page title or a character vector, and the function will return the graph of all the links beetween al the pages included in the vector
domain	The domain where is located the wiki
namespace	The namespace pages the function will graph

Value

An igraph object

Examples

```
# Graph of the 'Action' article in the french wiki.
graph_page('Action')

# Graphing a group of page
page <- c('Karl Marx', 'Classe sociale', 'Industrie')
# Return a graph where the 3 edges represents 'Karl Marx', 'Classe sociale' and 'Industrie',
# and the vertices the link present or not beetween this pages.
g <- graph_page(page)
```

page_anon	<i>Giving the proportion of anonymous contribution in a revisions list</i>
-----------	--

Description

Giving the proportion of anonmous contribution in a revisions list

Usage

```
page_anon(revisions)
```


Arguments

revisions Revision list built with page_revisions

Value

An integer.

See Also

Other page functions: [page_infos](#), [page_islink](#), [page_links](#), [page_ranking](#), [page_revisions](#)

Examples

```
# Downloading the list of the revisions of the article 'Action' in the french wiki
revisions <- page_revisions('Action')

# Returning the percentage of anonymous contribution in this article
page_anon(revisions)
```

page_cat

List all categories where the page is located

Description

List all categories where the page is located

Usage

```
page_cat(title, domain = "fr")
```

Arguments

title The title of the page
domain The domain where the wiki is located

Value

A character vector giving

Examples

```
# Downloading the list of categories in the action page
page_cat('Action')
```

page_infos *Getting some quantitative infos about a wiki page*

Description

Getting some quantitative infos about a wiki page

Usage

```
page_infos(title, domain = "fr")
```

Arguments

title	The title of the page
domain	The domain where the wiki is located

Value

A named list

See Also

Other page functions: [page_anon](#), [page_islink](#), [page_links](#), [page_ranking](#), [page_revisions](#)

Examples

```
# Getting a list of information for the 'Nationalism' page in the french wiki
page_infos('Nationalisme')
```

page_islink *Test if page(s) is(are) linked*

Description

Test if page(s) is(are) linked

Usage

```
page_islink(from, to, namespace = "0", domain = "fr")
```

Arguments

from	The title of the page from the link is supposed to go from
to	Either a string or a character vector representing the page(s) where the link(s) is(are) supposed to go to.
namespace	The namespace of the page to look.
domain	The domain where the wiki is located

Value

A character vector.

See Also

Other page functions: [page_anon](#), [page_infos](#), [page_links](#), [page_ranking](#), [page_revisions](#)

Examples

```
# Testing if the article 'Action' contains a link to 'Sociologie' in the french wiki.
page_islink('Action', 'Sociologie')
```

page_links

List of inter-articles links on a wiki page

Description

List of inter-articles links on a wiki page

Usage

```
page_links(x, domain = "fr", namespace = "0")
```

Arguments

x	The title of the page
domain	The domain of the wiki where the page is located
namespace	The namespace where the page need to be to be ept in the graph

Value

A character vector

See Also

Other page functions: [page_anon](#), [page_infos](#), [page_islink](#), [page_ranking](#), [page_revisions](#)

Examples

```
page_links('Action') # listing all the links contained in the 'action' article.
```

page_ranking	<i>Extracting the first contributors from a list of revisions</i>
--------------	---

Description

Extracting the first contributors from a list of revisions

Usage

```
page_ranking(revisions, thresold, keepAnon = TRUE)
```

Arguments

revisions	A list of revisions built with <code>page_revisions</code>
thresold	An integer, the function will return the thresold percent first contributors
keepAnon	A boolean indicate whether the anonymous should be kept into the list or not

Value

A character vector

See Also

Other page functions: [page_anon](#), [page_infos](#), [page_islink](#), [page_links](#), [page_revisions](#)

Examples

```
# Downloading the list of revisions
revisions <- page_revisions('Action')

# Extracting the list of the 10 percent first contributor of the list
page_ranking(revisions,10,TRUE)
```

page_revisions	<i>Downloading the list of contributions for one page</i>
----------------	---

Description

Downloading the list of contributions for one page

Usage

```
page_revisions(page, domain = "fr", contrib = F)
```

Arguments

page	The title of the page
domain	The domain where the wiki is located
contrib	A boolean indicating whether to return or not only contributors' names

Value

A data-frame containing the username of the user (or the IP if anonymous contribution), the timestamp, the size of the revision, a boolean indicating whether the contribution is anonymous or not, and the difference between the contribution and the previous

See Also

Other page functions: [page_anon](#), [page_infos](#), [page_islink](#), [page_links](#), [page_ranking](#)

Examples

```
# Downloading the list of contribution for the 'action' page in the french wiki
page_revisions('Action')
```

period	<i>Recode a date vector into periods</i>
--------	--

Description

Recode a date vector into periods

Usage

```
period(date, period = c("2005-05-01", "2009-12-01"))
```

Arguments

date	A timestamp vector
period	The breaks witch date should be cut with

Value

a numeric vector with the periods replacing the date

xpath	<i>Query a page by xpath means</i>
-------	------------------------------------

Description

Query a page by xpath means

Usage

```
xpath(url, xpath)
```

Arguments

url	URL of the page. Please note that this is different of the title of the wiki page.
xpath	XPATH query to execute

Value

A character vector

Index

[contrib_list](#), [2](#)
[corpus_contrib_create](#), [3](#), [4-7](#)
[corpus_contrib_data](#), [3](#), [4](#), [5-7](#)
[corpus_contrib_select](#), [3](#), [4](#), [5](#), [6](#), [7](#)
[corpus_page_create](#), [3-5](#), [6](#), [7](#)
[corpus_page_data](#), [3-6](#), [7](#)

[graph_page](#), [8](#)

[page_anon](#), [8](#), [10-13](#)
[page_cat](#), [9](#)
[page_infos](#), [9](#), [10](#), [11-13](#)
[page_islink](#), [9](#), [10](#), [10](#), [11-13](#)
[page_links](#), [9-11](#), [11](#), [12](#), [13](#)
[page_ranking](#), [9-11](#), [12](#), [13](#)
[page_revisions](#), [9-12](#), [12](#)
[period](#), [13](#)

[xpath](#), [14](#)