

Package ‘WikipediaR’

October 12, 2022

Type Package

Title R-Based Wikipedia Client

Version 1.1

Date 2016-02-05

Maintainer Avner Bar-Hen <Avner.Bar-Hen@mi.parisdescartes.fr>

Description Provides an interface to the Wikipedia web application programming interface (API), using internet connexion. Three functions provide details for a specific Wikipedia page : all links that are present, all pages that link to, all the contributions (revisions for main pages, and discussions for talk pages). Two functions provide details for a specific user : all contributions, and general information (as name, gender, rights or groups). It provides additional information compared to others packages, as WikipediR. It does not need login. The multiplex network that can be constructed from the results of the functions of WikipediaR can be modeled as Stochastic Block Model as in Barbillon P., Donnet, S., Lazega E., and Bar-Hen A. : Stochastic Block Models for Multiplex networks: an application to networks of researchers, ArXiv 1501.06444, <http://arxiv.org/abs/1501.06444>.

License GPL (> 2)

Depends R (>= 2.10.0), XML (>= 2.6.3), httr (>= 1.0.0)

URL <http://www.r-project.org>

Repository CRAN

RoxygenNote 5.0.1

NeedsCompilation no

Author Avner Bar-Hen [aut, cre],
Louise Baschet [ctb],
Francois-Xavier Jollois [ctb],
Jeremie Riou [ctb]

Date/Publication 2016-02-05 19:19:00

R topics documented:

WikipediaR-package	2
backLinks	4
contribs	6
links	8
print.backLinksClass	9
print.contribsClass	10
print.linksClass	10
print.userContribsClass	11
print.userInfoClass	12
testWikiPage	12
testWikiUser	13
userContribs	14
userInfo	16

Index	18
--------------	-----------

WikipediaR-package	<i>R-Based Wikipedia Client</i>
--------------------	---------------------------------

Description

Provides an interface to the Wikipedia web API.

Three functions provide details for a specific Wikipedia page ; the `links` function lists all links that are present, the `backLinks` function lists all pages that link to, the `contribs` function lists all the contributions (revisions for main pages, and discussions for talk pages). The page can be defined by the parameter "page", as the title, i.e. a character string, or the page ID, a numeric value. The character string title can include spaces and special characters, and lower/upper case letters are taking in account.

Two functions provides details for a specific user ; the `userContribs` function lists all contributions, and the `userInfo` function provides general information (as name, gender, rights or groups). The user is defined by his or her name. Lower/upper case letters are taking in account.

The domain can be specified in all the functions, by the parameter "domain". the default domain is "en" for the <https://en.wikipedia.org>.

It provides additional information compared to others packages, as WikipediR. It does not need login.

The multiplex network that can be constructed from the results of the functions of WikipediaR can be modeled as Stochastic Block Model as in Barbillon P. et al.

Details

Package:	WikipediaR
Type:	Package
Version:	1.1
Date:	2016-01-25

```

License:      GPL (>2)
Depends:     R (>= 2.10.0), XML (>=2.6.3), httr (>=1.0.0)
URL:         http://www.r-project.org
Repository:  CRAN
Packaged:    2016-02-05 09:57:22 UTC; Louise
Built:       R 3.2.3; ; 2016-02-05 09:57:22 UTC; windows

```

As part of a PRES Sorbonne Paris Cite project, Paris Descartes statisticians, computer scientists and sociologists Paris Diderot Sciences Po are working on the problem of multi-level networks. One part of the project is to analyze data extracted from Wikipedia with the free software R.

Like the `twitter` package that provides an interface to the Twitter web API, the objective of the `WikipediaR` package is to provide a way to access to data extracted from Wikipedia, and return it in a exploitable format in R. API means application programming interface.

A package already exists with a similar objective : `WikipediR`. This package is under progress, and as it permits modifications on the Wikipedia base, it needs a connexion, with rights. To see more details about this package, you can visit <http://ironholds.org/blog/introducing-wikipedir/>.

Our package use the `XML` package to interact with wikipedia, via MediaWiki API syntax. You can see this syntax here: <http://en.wikipedia.org/w/api.php>.

Which others packages interact with MediaWiki API ? The Tiki Wiki CMS/Groupware framework has an R plugin (`PluginR`) to run R code from wiki pages, and use data from their own collected web databases (trackers). A demo: <http://r.tiki.org>

The `wikibooks` package provides functions and datasets of the german WikiBook "GNU R".

Remark1: "fr" and "en" domains have been tested, but others can lead to not anticipated problem. Trying domain="gu" is at your peril... The encoding is UTF-8 for most of the output.

Remark2: as the functions get informations on internet in real time, the execution time depends on your internet connection !

Index:

<code>backLinks</code>	lists pages that link to the Wikipedia page
<code>contribs</code>	lists contributions for a specific wikipedia page
<code>links</code>	links on a Wikipedia page
<code>testWikiPage</code>	internal function <code>testWikiPage</code>
<code>testWikiUser</code>	internal function <code>testWikiPage</code>
<code>userContribs</code>	list of contributions for a specific user
<code>userInfo</code>	General information for a Wikipedia user

Author(s)

Avner Bar-Hen <Avner.Bar-Hen@mi.parisdescartes.fr>, with contributions of Louise Baschet, Francois-Xavier Jollois, Jeremie Riou

Maintainer: Avner Bar-Hen <Avner.Bar-Hen@mi.parisdescartes.fr>

References

Barbillon P., Donnet, S., Lazega E., and Bar-Hen A. : Stochastic Block Models for Multiplex networks: an application to networks of researchers, ArXiv 1501.06444, <http://arxiv.org/abs/1501.06444>.

See Also

[WikipediR](#)

Examples

```
## Not run:
## wikipedia links and external links that are present
links("Louis Pasteur") # default domain : en
## if the specified page needs to be redirected
links(page = "Structure baschet", domain = "fr")
# warning message

## wikipedia pages that link to the page about Luxor Temple
bl.Luxor <- backLinks(domain = "en", page = "Luxor Temple")
# in how many main pages and discussions this page is linked ?
table(bl.Luxor$backLinks$nsct)

## contributions on a page specified by its page ID
contrib.5636 <- contribs(5636, domain = "en")
# get the last non anonymous contributor
last.contrib.5636 <- contrib.5636$contribs[which(contrib.5636$contribs$userid!=0),]$user[1]
last.contrib.5636

## number of contributions for this user
nrow(userContribs(user.name = last.contrib.5636, domain = "en")$contribs)
## and informations about this user
userInfo(user.name = last.contrib.5636, domain = "en",
  usprop="groups|rights|editcount|gender" )

## End(Not run)
```

backLinks

lists pages that link to the Wikipedia page

Description

lists all pages that link to a specific Wikipedia page.

Usage

```
backLinks(page = NULL, domain = "en")
```

Arguments

page	numeric identifier or character title of the specific wikipedia page
domain	a character value specifying the language of the wikipedia page. The default value is "en" for "english language".

Details

This function uses API query syntax: "list=backlinks". For more details, see <https://www.mediawiki.org/wiki/API:Backlinks>

Value

an object of class backLinksClass containing:

- call the command line
- page title, identification number and domain of the Wikipedia page, created only if the page exists.
- backLinks a data frame containing title, the titles of the pages that link to the specific page, pageid, the pages IDs, ns, numbers of the namespaces (identification of the type of pages, defining nscat et nssubj, the two next variables), nscat, the categories of linked pages (Subject, Talk, or Virtual), nssubj, the subjects of the linked pages (Main Article, User, Wikipedia, File, MediaWiki, Template, Help, Category, Protal, Book, Draft, Education Program, TimedText, Module, Topic, Special, Media, Other). For more details about namespace, see http://en.wikipedia.org/wiki/Wikipedia:Namespace#Subject_namespaces If the page has no back link or does not exist, this item is not created.
- testWikiPage a list of four elements,
 - takeOnlyFirst a boolean indicating if the class of page parameter is invalid, for example vector, list, matrix..., and in that case, only the first element is considered.
 - redirPage title of the redirected page. This item is NULL if the page is not redirected.
 - test an integer with value:
 - * 4 for invalid domain,
 - * 3 for an empty parameter page,
 - * 2 when Wikipedia does not have an article with this exact name,
 - * 1 for ambiguous page, direct or redirect,
 - * 0 for valid an unambiguous page, direct or redirect.
 - warnMessage is a vector of warning messages.

Author(s)

Avner Bar-Hen, Louise Baschet, Francois-Xavier Jollois, Jeremie Riou

See Also

print.backLinksClass links

Examples

```
## Not run:
#' # a simple example
backLinks("Louis Pasteur")

backLinks.Baschet <- backLinks(page = "Cristal Baschet", domain = "fr")
table(backLinks.Baschet$backLinks$nscat)

## example with no back link
backLinks(page = 976, domain = "en" )

# with a page that not exist (at the moment of the redaction of this help page)
backLinks("zzzzz")

## End(Not run)
```

contribs

lists contributions of a specific wikipedia page

Description

provide the list of the contributions of a wikipedia page: discussions for Talk pages, or revisions for Subject pages.

Usage

```
contribs(page = NULL, domain = "en", rvprop = "user|userid|timestamp")
```

Arguments

page	numeric identifier or character title of the specific wikipedia page
domain	a character value specifying the language of the wikipedia page. The default value is "en" for "english language".
rvprop	Which properties to get for each revision (separate with ' '):

- ids: the ID of the revision
- flags: revision flags (minor)
- timestamp: the timestamp of the revision, i.e. day and time
- user: user that made the revision
- userid: user id of revision creator
- size: length (bytes) of the revision
- sha1: SHA-1 (base 16) of the revision
- contentmodel: content model id
- comment: comment by the user for revision
- parsedcomment: parsed comment by the user for the revision
- content: text of the revision
- tags: tags for the revision

Default: user|userid|timestamp

Details

This function uses the API query syntax: "prop=revisions". For more details, see <https://www.mediawiki.org/wiki/API:Backlinks> If the page is ambiguous, the contributions correspond to the modification on this ambiguous page, not on the possible linked pages. If the user is anonymous, the user in the output corresponds to the IP address, and the userid is zero.

Value

an object of class contribsClass:

- call the command line
- page title and identification number of the Wikipedia page
- contribs a data frame containing asked properties of the contribs, by default : user, userid, timestamp. If the user does not exist or has no contribution, this item is not created.
- testWikiPage a list of four elements,
 - takeOnlyFirst a boolean indicating if the class of page parameter is invalid, for example vector, list, matrix..., and in that case, only the first element is considered.
 - redirPage title of the redirected page. This item is NULL if the page is not redirected.
 - test an integer with value:
 - * 4 for invalid domain,
 - * 3 for an empty parameter page,
 - * 2 when Wikipedia does not have an article with this exact name,
 - * 1 for ambiguous page, direct or redirect,
 - * 0 for valid an unambiguous page, direct or redirect.
 - warnMessage is a vector of warning messages.

Author(s)

Avner Bar-Hen, Louise Baschet, Francois-Xavier Jollois, Jeremie Riou

See Also

print.contribsClass userInfo

Examples

```
## Not run:
## numeric page identifier as parameter
contribs(domain = "fr", page = 108907)

## title page as parameter
contribs(domain = "en", page = "Wikipedia")

## End(Not run)
```

links	<i>lists links on a Wikipedia page</i>
-------	--

Description

lists all links (to wikipedia and to external url) that are present in a specific wikipedia page.

Usage

```
links(page = NULL, domain = "en")
```

Arguments

page	numeric identifier or character title of the specific wikipedia page
domain	a character value specifying the language of the wikipedia url. The default value is "en" for "english language".

Details

This function uses the MediaWiki API query syntax: "prop=links". For more details, see http://www.mediawiki.org/wiki/API:Properties#links_.2F_pl.

Value

an object of class linksClass containing:

- call the command line
- page title, identification number, and domain of the Wikipedia page
- links a data frame containing title, the titles of the linked pages, ns, numbers of the namespaces (identification of the type of pages, defining nscat et nssubj, the two next variables), nscat, the categories of linked pages (Subject, Talk, or Virtual), nssubj, the subjects of the linked pages (Main Article, User, Wikipedia, File, MediaWiki, Template, Help, Category, Portal, Book, Draft, Education Program, TimedText, Module, Topic, Special, Media, Other). For more details about namespace, see <http://www.mediawiki.org/wiki/Help:Namespaces> and http://en.wikipedia.org/wiki/Wikipedia:Namespace#Subject_namespaces
- extlinks a vector containing the list of url of external links. If there is no external link in the page, this item is not created.
- testWikiPagea list of four elements,
 - takeOnlyFirst a boolean indicating if the class of page parameter is invalid, for example vector, list, matrix..., and in that case, only the first element is considered.
 - redirPage title of the redirected url. This item is NULL if the page is not redirected.
 - test an integer with value:
 - * 4 for invalid domain,
 - * 3 for an empty parameter page,
 - * 2 when Wikipedia does not have an article with this exact name,

- * 1 for ambiguous page, direct or redirect,
- * 0 for valid an unambiguous page, direct or redirect.
- warnMessage is a vector of warning messages.

Author(s)

Avner Bar-Hen, Louise Baschet, Francois-Xavier Jollois, Jeremie Riou

See Also

`print.linksClass backLinks`

Examples

```
## Not run:  
# a simple example  
links("Louis Pasteur") # default domain : en  
  
# with a redirected page  
links.Obama <- links(page ="Obama")  
links.Obama  
# warning message  
  
# a simple example with page specified by its page ID number  
links(page = 976, domain = "fr" )  
  
# with a page that not exist (at the moment of the redaction of this help page)  
links("zzzzz")  
  
## End(Not run)
```

`print.backLinksClass` *S3method print backLinks*

Description

S3method to print backLinksClass objects, results of the backLinks function

Usage

```
## S3 method for class 'backLinksClass'  
print(x, maxprint = 10, ...)
```

Arguments

<code>x</code>	the backLinksClass object, result of the backLinks function
<code>maxprint</code>	the maximal number of pages to print
<code>...</code>	others parameters

Author(s)

Avner Bar-Hen, Louise Baschet, Francois-Xavier Jollois, Jeremie Riou

See Also

backLinks

`print.contribsClass` *S3method print contribs*

Description

S3method to print contribsClass objects, results of the contribs function

Usage

```
## S3 method for class 'contribsClass'  
print(x, maxprint = 10, ...)
```

Arguments

<code>x</code>	the contribsClass object, result of the contribs function
<code>maxprint</code>	the maximal number of pages to print
<code>...</code>	others parameters

Author(s)

Avner Bar-Hen, Louise Baschet, Francois-Xavier Jollois, Jeremie Riou

See Also

contribs

`print.linksClass` *S3method print links*

Description

S3method to print linksClass objects, results of the links function

Usage

```
## S3 method for class 'linksClass'  
print(x, maxprint = 10, ...)
```

Arguments

x	the linksClass object, result of the links function
maxprint	the maximal number of pages to print
...	others parameters

Author(s)

Avner Bar-Hen, Louise Baschet, Francois-Xavier Jollois, Jeremie Riou

See Also

links

`print.userContribsClass`
S3method print userContribs

Description

S3method to print userContribsClass objects, results of the userContribs function

Usage

```
## S3 method for class 'userContribsClass'  
print(x, maxprint = 10, ...)
```

Arguments

x	the userContribsClass object, result of the userContribs function
maxprint	the maximal number of contributions to print
...	others parameters

Author(s)

Avner Bar-Hen, Louise Baschet, Francois-Xavier Jollois, Jeremie Riou

See Also

userContribs

print.userInfoClass *S3method print userInfo*

Description

S3method to print userInfoClass objects, results of the userInfo function

Usage

```
## S3 method for class 'userInfoClass'  
print(x, ...)
```

Arguments

x	the userInfoClass object, result of the userInfo function
...	others parameters

Author(s)

Avner Bar-Hen, Louise Baschet, Francois-Xavier Jollois, Jeremie Riou

See Also

userInfo

testWikiPage *internal function testWikiPage*

Description

internal function to test validity of a wikidepia page, called in links, backLinks, and contribs functions.

Usage

```
testWikiPage(page = NULL, domain = "en")
```

Arguments

page	numeric identifier or character title of the specific wikipedia page
domain	a character value providing the language of the wikipedia page. The default value is "en" for "english language".

Value

a list of four elements, the first is `takeOnlyFirst`, a boolean indicating if the class of page parameter is invalid, for example `vector`, `list`, `matrix...`, and in that case, the only the first element is considered. The second element is `redirPage`, the title page of the redirected page. The third element is `test`, an integer with value:

- 4 for invalid domain,
- 3 for an empty parameter page,
- 2 when Wikipedia does not have an article with this exact name,
- 1 for ambiguous page, direct or redirect,
- 0 for valid an unambiguous page, direct or redirect.

The last element, `warnMessage`, is a vector of warning messages.

Author(s)

Avner Bar-Hen, Louise Baschet, Francois-Xavier Jollois, Jeremie Riou

<code>testWikiUser</code>	<i>internal function testWikiUser</i>
---------------------------	---------------------------------------

Description

internal function to test validity of a wikidepia domain and user, called in `userInfo` and `userContribs` functions.

Usage

```
testWikiUser(user.name = NULL, domain = "en")
```

Arguments

<code>user.name</code>	a character value providing the name of the specific wikipedia user
<code>domain</code>	a character value providing the language of the wikipedia page. The default value is "en" for "english language".

Value

a list of four elements, the first is `takeOnlyFirst`, a boolean indicating if the class of `user.name` parameter is invalid, for example `vector`, `list`, `matrix...`, and in that case, the only the first element is considered. The second element is `test`, an integer with value:

- 4 for invalid domain,
- 3 for an empty parameter `user.name`,
- 2 when Wikipedia does not have an user with this exact name,
- 0 for valid existing `user.name`.

The third element is `user`, a vector containing the user name, the user identifier, and domain. If the user does not exist, this item is not created. The last element, `warnMessage`, is a vector of warning messages.

Author(s)

Avner Bar-Hen, Louise Baschet, Francois-Xavier Jollois, Jeremie Riou

userContribs	<i>lists contributions for a specific user</i>
--------------	--

Description

lists contributions for a specific user: discussions for Talk pages, and revisions for Subject pages.

Usage

```
userContribs(user.name = NULL, domain = "en",
             ucprop = "ids|title|timestamp|comment|sizediff|flags")
```

Arguments

<code>user.name</code>	a character value providing the name of the user
<code>domain</code>	a character value providing the language of the wikipedia page. The default value is "en" for "english language".
<code>ucprop</code>	<p>Include informations (separate with ' ') :</p> <ul style="list-style-type: none"> • <code>ids</code> the page ID and revision ID • <code>title</code> the title and namespace ID of the page. This parameter is not optional. • <code>timestamp</code> the timestamp of the edit, i.e. day and time • <code>comment</code> the comment of the edit • <code>parsedcomment</code> the parsed comment of the edit • <code>size</code> the new size of the edit • <code>sizediff</code> the size delta of the edit against its parent • <code>flags</code> flags of the edit • <code>patrolled</code> patrolled edits • <code>tags</code> tags for the edit

Default: `ids|title|timestamp|comment|sizediff|flags`

Details

This function uses the API query syntax: `"list=usercontribs"`. For more details, see <https://www.mediawiki.org/wiki/API:Usercontribs>. Additionally to the titles of the modified pages, this function always returns in the `contribs` item the `ns` for namespace (identification of the type of pages), the `nscat` for the category of the pages (Subject, Talk, or Virtual), and the subject of the pages.

Value

an object of class `userContribsClass`:

- call the command line
- user a vector containing the user name and the user identifier
- contribs a data frame containing the asked properties of the revisions. If `ids` is asked, the corresponding results are `pageid`, `revid`, and `parentid`. Three others informations are automatically added : `ns`, numbers of the namespaces (identification of the type of pages, defining `nscat` et `nssubj`, the two next variables), `nscat`, the categories of linked pages (Subject, Talk, or Virtual), `nssubj`, the subjects of the linked pages (Main Article, User, Wikipedia, File, MediaWiki, Template, Help, Category, Protal, Book, Draft, Education Program, TimedText, Module, Topic, Special, Media, Other) . This item is provided if the user has at least one contribution. For more details about namespace, see <http://www.mediawiki.org/wiki/Help:Namespaces> and http://en.wikipedia.org/wiki/Wikipedia:Namespace#Subject_namespaces
- testWikiUser A list of three elements. The first is `takeOnlyFirst`, a boolean indicating if the class of `user.name` parameter is invalid, for example `vector`, `list`, `matrix...`, and in that case, the only the first element is considered. The second element is `test`, an integer with value:
 - 4 for invalid domain,
 - 3 for an empty parameter user,
 - 2 when Wikipedia does not have an user with this exact name,
 - 0 for valid existing user.

The last element, `warnMessage`, is a vector of warning messages.

Author(s)

Avner Bar-Hen, Louise Baschet, Francois-Xavier Jollois, Jeremie Riou

See Also

`userInfo`

Examples

```
## Not run:
LouiseContribs <- userContribs(user.name = "Louise", domain = "en")
## try a user that does not exist (at the moment of the redaction of this help page)
userContribs(user.name="Louise Baschet", domain ="fr")

## End(Not run)
```

userInfo *provides general information about a Wikipedia user*

Description

provides general information about a specific Wikipedia user

Usage

```
userInfo(user.name = NULL, domain = "en",
         usprop = "groups|implicitgroups|rights|editcount|registration|emailable|gender")
```

Arguments

user.name	a character value for the name of a user
domain	a character value providing the language of the wikipedia page. The default value is "en" for "english language".
usprop	<p>What pieces of information to include (separate with ' '):</p> <ul style="list-style-type: none"> • groups: lists all the groups the user(s) belongs to • implicitgroups: lists all the groups a user is automatically a member of • rights: lists all the rights the user(s) has • editcount: adds the user's edit count • registration: adds the user's registration timestamp • emailable: tags if the user can and wants to receive email through [[Special:Emailuser]] • gender: tags the gender of the user. Returns "male", "female", or "unknown" <p>Default: groups implicitgroups rights editcount registration emailable gender</p>

Details

This function uses the MediaWiki API query syntax: "list=users". For more details, see <https://www.mediawiki.org/wiki/API:Users>

Value

an object of class userInfoClass containing:

- call the command line
- rights only if asked: a vector of rights. If not asked, this item does not exist.
- groups only if asked: a vector of groups names. If not asked, this item does not exist.
- implicitgroups only if asked: a vector of implicit groups names. If not asked, this item does not exist.
- info a data frame containing the other asked properties of user (at least userid and name).

- testWikiUser A list of three elements. The first is takeOnlyFirst, a boolean indicating if the class of user.name parameter is invalid, for example vector, list, matrix..., and in that case, the only the first element is considered. The second element is test, an integer with value:
 - 4 for invalid domain,
 - 3 for an empty parameter user,
 - 2 when Wikipedia does not have an user with this exact name,
 - 0 for valid existing user.

The last element, warnMessage, is a vector of warning messages.

Author(s)

Avner Bar-Hen, Louise Baschet, Francois-Xavier Jollois, Jeremie Riou

Examples

```
## Not run:
LouiseInfo <- userContribs(user.name = "Louise", domain = "en")
LouiseInfo
bobInfo <- userInfo(user.name = "bob", domain = "en")
bobInfo
## try a user that does not exist (at the moment of the redaction of this help page)
userInfo(user.name="Louise Baschet", domain ="fr")

## End(Not run)
```

Index

* package

WikipediaR-package, [2](#)

[backLinks](#), [4](#)

[contribs](#), [6](#)

[links](#), [8](#)

[print.backLinksClass](#), [9](#)

[print.contribsClass](#), [10](#)

[print.linksClass](#), [10](#)

[print.userContribsClass](#), [11](#)

[print.userInfoClass](#), [12](#)

[testWikiPage](#), [12](#)

[testWikiUser](#), [13](#)

[userContribs](#), [14](#)

[userInfo](#), [16](#)

[WikipediaR \(WikipediaR-package\)](#), [2](#)

[WikipediaR-package](#), [2](#)

[WikipediR](#), [4](#)