

Package ‘wikipediatrend’

August 29, 2016

Title Public Subject Attention via Wikipedia Page View Statistics

Version 1.1.10

Date 2016-03-20

Author Peter Meissner [aut, cre], R Core Team [ctb]

Maintainer Peter Meissner <retep.meissner@gmail.com>

Description Public attention is an interesting field of study. The internet not only allows to access information in no time on virtually any subject but via page access statistics gathered by website authors the subject of attention as well can be studied. For the omnipresent Wikipedia those access statistics are made available via 'http://stats.grok.se' a server providing the information as file dumps as well as as web API. This package provides an easy to use, consistent and traffic minimizing approach to make those data accessible within R.

License GPL (>= 2)

Depends R (>= 3.0.0)

Imports jsonlite (>= 0.9.13), stringr (>= 0.6.2), rvest (>= 0.2.0), httr (>= 0.6.1), utils, xml2 (>= 0.1.2), hellno (>= 0.0.1)

Suggests testthat, knitr, ggplot2, devtools, dplyr, magrittr, AnomalyDetection, BreakoutDetection, cranlogs

BugReports <https://github.com/petermeissner/wikipediatrend/issues>

VignetteBuilder knitr

Additional_repositories <http://ghrr.github.io/drat>

RoxygenNote 5.0.1

NeedsCompilation no

Repository CRAN

Date/Publication 2016-03-20 16:40:50

R topics documented:

html	2
html2	3
prev_month_end	3
prev_month_start	3
print.wp_df	4
read_utf8_csv	4
toUTF8	4
wp_add_to_cache	5
wp_cache	5
wp_cache_default	5
wp_cache_file	6
wp_cache_load	6
wp_cache_reset	6
wp_check_date_inputs	7
wp_date	7
wp_day	8
wp_download_data	9
wp_expand_ts	10
wp_get_cache	10
wp_get_data	10
wp_http_header	11
wp_jsons_to_df	11
wp_linked_pages	11
wp_load	12
wp_month	12
wp_prepare_urls	13
wp_save	13
wp_save_cache	14
wp_set_cache_file	14
wp_set_print_options	14
wp_trend	15
wp_wday	16
wp_year	16
wp_yearmonth	17
write_utf8_csv	17
Index	18

html

*function downloading content***Description**

function downloading content

Usage

```
html(url, ...)
```

Arguments

url	url to content
...	further arguments passed trough to http::GET()

html2	<i>function downloading content</i>
-------	-------------------------------------

Description

function downloading content

Usage

```
html2(url, ...)
```

Arguments

url	url to content
...	further arguments passed trough to http::GET()

prev_month_end	<i>function to get date of end of prev month</i>
----------------	--------------------------------------------------

Description

function to get date of end of prev month

Usage

```
prev_month_end()
```

prev_month_start	<i>function to get date of start of prev month</i>
------------------	----------------------------------------------------

Description

function to get date of start of prev month

Usage

```
prev_month_start()
```

<code>print.wp_df</code>	<i>a wikipediastrend specific data frame</i>
--------------------------	----------------------------------------------

Description

a wikipediastrend specific data frame

Usage

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

Arguments

<code>x</code>	the thing to be printed
<code>...</code>	print.default parameters

<code>read_utf8_csv</code>	<i>function to read csv file with UTF-8 characters (even under Windwos) that were created by write_U</i>
----------------------------	----------------------------------------------------------------------------------------------------------

Description

function to read csv file with UTF-8 characters (even under Windwos) that were created by write_U

Usage

```
read_utf8_csv(file)
```

Arguments

<code>file</code>	file name / path where to get the data
-------------------	----------------------------------------

<code>toUTF8</code>	<i>function to convert character vectors to UTF-8 encoding</i>
---------------------	----------------------------------------------------------------

Description

function to convert character vectors to UTF-8 encoding

Usage

```
toUTF8(x)
```

Arguments

<code>x</code>	the vector to be converted
----------------	----------------------------

wp_add_to_cache	<i>function adding downloaded data to the cache</i>
-----------------	-----------------------------------------------------

Description

function adding downloaded data to the cache

Usage

```
wp_add_to_cache(df)
```

Arguments

df	data frame to be added to cache
----	---------------------------------

wp_cache	<i>a cache for downloads</i>
----------	------------------------------

Description

a cache for downloads

Usage

```
wp_cache
```

Format

An object of class environment of length 2.

wp_cache_default	<i>function for building default cache file name</i>
------------------	------------------------------------------------------

Description

function for building default cache file name

Usage

```
wp_cache_default()
```

wp_cache_file	<i>function for determining where downloaded information can be stored</i>
---------------	----------------------------------------------------------------------------

Description

function for determining where downloaded information can be stored

Usage

```
wp_cache_file()
```

wp_cache_load	<i>function for loading cache on disk into memory</i>
---------------	-------------------------------------------------------

Description

function for loading cache on disk into memory

Usage

```
wp_cache_load(file = wp_cache_file())
```

Arguments

file	where to find previously saved cache
------	--------------------------------------

wp_cache_reset	<i>function for resetting cache/cachefile as if package was just loaded</i>
----------------	-----------------------------------------------------------------------------

Description

function for resetting cache/cachefile as if package was just loaded

Usage

```
wp_cache_reset()
```

wp_check_date_inputs *Helper function for wp_trend()*

Description

Function that checks if the time span given by from and to (passed down from wp_trend) are complying with logical constraints: from not prior to first available data; to not past today; to not prior to from

Usage

```
wp_check_date_inputs(from, to)
```

Arguments

from	first date of timespan to check
to	second date of timespan to check

wp_date *Package specific 'as.Date()' function*

Description

wp_date() provides a generic function and several methods to transform something into a date, just like as.Date() from the base package. While most methods were copied 1:1 from as.Date() some changes were made that alter the usual behavior: (1) Whenever wp_date() is applied to an empty string "" or any other not-standard-unambiguous format the function will return NA and throw out a warning instead of simply kill itself with an error. (2) Whenever, as.Date usually is in the need of an origin supplied and does not get one wp_date() will assume the origin to be "1970-01-01" and throw an error instead of kill itself with an error. All these changes are made with the special purpose to make wp_trend() also with dirty, unreasonable data (e.g. "" or "2012-13-01" or "2012-02-32") and are not recommended for any other than wikipediatrend internal use.

Usage

```
wp_date(x, ...)  
  
## S3 method for class 'character'  
wp_date(x, format = "", ...)  
  
## Default S3 method:  
wp_date(x, ...)  
  
## S3 method for class 'numeric'  
wp_date(x, origin, ...)
```

```
## S3 method for class 'dates'
wp_date(x, ...)

## S3 method for class 'factor'
wp_date(x, ...)

## S3 method for class 'POSIXct'
wp_date(x, tz = "UTC", ...)

## S3 method for class 'POSIXlt'
wp_date(x, ...)
```

Arguments

x	See ?as.Date
...	See ?as.Date
format	See ?as.Date
origin	See ?as.Date
tz	See ?as.Date

Methods (by class)

- character: Difference to as.Date.character(): Failing to parse the date, wp_date() will give return a warning and NA instead of killing itself with an error.
- default: same as as.Date()
- numeric: Difference to as.Date.character(): Whenever an origin is needed but not supplied instead of breaking, wp_date() will raise a warning and proceed by assuming "1970-01-01" to be the origin instead of killing itself with an error.
- dates: same as as.Date()
- factor: same as as.Date()
- POSIXct: same as as.Date()
- POSIXlt: using as.Date()

wp_day

Function to extract the day

Description

Function to extract the day from a timestamp of e.g. class POSIXlt, POSIXct, Date, or character. If the option is of type character it should be in the form of yyyy-mm-dd.

Usage

```
wp_day(timestamp)

## S3 method for class 'POSIXlt'
wp_day(timestamp)

## Default S3 method:
wp_day(timestamp)
```

Arguments

timestamp Some sort of timestamp e.g. of class POSIXlt, POSIXct, Date, or character. If the option is of type character it should be in the form of yyyy-mm-dd.

Methods (by class)

- POSIXlt: extract day from timestamp
- default: extract day from timestamp

wp_download_data *function downloading prepared URLs*

Description

function downloading prepared URLs

Usage

```
wp_download_data(urls, wait = 1)
```

Arguments

urls a vector of urls to be downloaded

wait the time to wait in seconds before downloading the next chunk (default=1)

wp_expand_ts	<i>Helper function for wp_trend that transforms a time span into a series of months</i>
--------------	-----------------------------------------------------------------------------------------

Description

Helper function for wp_trend that transforms a time span into a series of months

Usage

```
wp_expand_ts(from, to, by)
```

Arguments

from	start of timespan
to	end of timespan
by	in which time unit expansion should take place: "day", "month", "year"

wp_get_cache	<i>function to get cache content</i>
--------------	--------------------------------------

Description

function to get cache content

Usage

```
wp_get_cache()
```

wp_get_data	<i>function for getting data (download + extraction)</i>
-------------	----------------------------------------------------------

Description

function for getting data (download + extraction)

Usage

```
wp_get_data(urls)
```

Arguments

urls	urls to be downloaded
------	-----------------------

wp_http_header	<i>function for pasting together the HTTP user-agent field</i>
----------------	----------------------------------------------------------------

Description

function for pasting together the HTTP user-agent field

Usage

```
wp_http_header()
```

wp_jsons_to_df	<i>Helper function for wp_trend</i>
----------------	-------------------------------------

Description

The function does accepts a list of JSON snippets and transforms the information into a data frame

Usage

```
wp_jsons_to_df(wp_json, page)
```

Arguments

wp_json	The results from server request in JSON format
page	the name of the downloaded page

wp_linked_pages	<i>function looking for other languages of a page</i>
-----------------	-------------------------------------------------------

Description

function looking for other languages of a page

Usage

```
wp_linked_pages(page, lang)
```

Arguments

page	title of the wikipedia article to look for links to other languages
lang	language (shorthand, e.g. "en" for English or "de" for German and so on) for the afor given title

wp_load	<i>Helper function for wp_trend that loads previous saved wp_trend results</i>
---------	--------------------------------------------------------------------------------

Description

Helper function for wp_trend that loads previous saved wp_trend results

Usage

```
wp_load(file = wp_cache_file())
```

Arguments

file	name of the file from which previous wp_trend() results should be loaded
------	--------------------------------------------------------------------------

wp_month	<i>Function to extract the month</i>
----------	--------------------------------------

Description

Function to extract the month from a timestamp of e.g. class POSIXlt, POSIXct, Date, or character. If the option is of type character it should be in the form of yyyy-mm-dd.

Usage

```
wp_month(timestamp)

## S3 method for class 'POSIXlt'
wp_month(timestamp)

## Default S3 method:
wp_month(timestamp)
```

Arguments

timestamp	Some sort of timestamp e.g. of class POSIXlt, POSIXct, Date, or character. If the option is of type character it should be in the form of yyyy-mm-dd.
-----------	-------------------------------------------------------------------------------------------------------------------------------------------------------

Methods (by class)

- POSIXlt: extract @export
- default: extract @export

wp_prepare_urls	<i>function preparing URLs for download</i>
-----------------	---------------------------------------------

Description

function preparing URLs for download

Usage

```
wp_prepare_urls(page, from, to, lang)
```

Arguments

page	page title supplied by wp_trend()
from	from supplied by wp_trend()
to	to supplied by wp_trend()
lang	lang supplied by wp_trend()

wp_save	<i>Helper function for wp_trend</i>
---------	-------------------------------------

Description

Function writes wp_trend() data to a CSV file that serves as cache/storage

Usage

```
wp_save(res, file = wp_cache_file())
```

Arguments

res	data retrieved by wp_trend()
file	where to save/cache results (defaults to wp_cache_file())

wp_save_cache *functio to save cache to file*

Description

functio to save cache to file

Usage

```
wp_save_cache(file = wp_cache_file())
```

Arguments

file where to store in memory cache (e.g. "myData.csv")

wp_set_cache_file *function to set a different cachefile as session default*

Description

function to set a different cachefile as session default

Usage

```
wp_set_cache_file(file)
```

Arguments

file name of the file to be used as cache

wp_set_print_options *function for setting print options for print.wp_df()*

Description

function for setting print options for print.wp_df()

Usage

```
wp_set_print_options(x)
```

Arguments

x a list of options, e.g. list(rows=35, width=50) or list(rows=Inf, width=Inf)

wp_trend

Function for getting access statistics for wikipedia pages

Description

Function for getting access statistics for wikipedia pages

Usage

```
wp_trend(page, from = prev_month_start(), to = prev_month_end(),
         lang = "en", file = "")
```

Arguments

page	The name of the Wikipedia page as to be found in the URL to the wikipedia article. If e.g. the URL is: https://en.wikipedia.org/wiki/Peter_Fox_(musician) , than the page name equals to Peter_Fox_(musician).
from	The starting date of the timespan for which access statistics should be retrieved - note that there is no data prior to 2007-12-01. Supply some sort of timestamp e.g. of class POSIXlt, POSIXct, Date, or character. If the option is of type character it should be in the form of yyyy-mm-dd.
to	The last date for which access statistics should be retrieved. Supply some sort of timestamp e.g. of class POSIXlt, POSIXct, Date, or character. If the option is of type character it should be in the form of yyyy-mm-dd.
lang	The language shorthand identifying which Wikipedia access statistics are to be used: e.g. "en" for the English version found at https://en.wikipedia.org , "de" for the German version found at https://de.wikipedia.org or perhaps "als" for the Alemannic dialect found under https://als.wikipedia.org/ .
file	Where to store data retrieved by function call? The data is stored in CSV format. If an already existing file is used for storage, the old data will not be deleted but instead new data will be added to this file! The cache in memory before the function call will be saved and restored afterwards. If you prefer automatic mirroring of in memory cache and cache on disk

Examples

```
library(wikipediatrend)
wp_trend( page = c("Cheese", "K\u00e4se"),
         from = "2014-11-01",
         to   = "2014-11-30",
         lang = c("en", "de"),
         file = wp_cache_file()
       )
```

wp_wday *Function to extract the day of the week*

Description

Function to extract the day from a timestamp of e.g. class POSIXlt, POSIXct, Date, or character. If the option is of type character it should be in the form of yyyy-mm-dd.

Usage

```
wp_wday(timestamp, startmonday = T)

## S3 method for class 'POSIXlt'
wp_wday(timestamp, startmonday = T)

## Default S3 method:
wp_wday(timestamp, startmonday = T)
```

Arguments

timestamp	Some sort of timestamp e.g. of class POSIXlt, POSIXct, Date, or character. If the option is of type character it should be in the form of yyyy-mm-dd.
startmonday	whether the week should start on Monday (TRUE, the default) or it should start on Sunday (FALSE)

Methods (by class)

- POSIXlt: method for POSIXlt
- default: method for everything coercable by as.POSIXlt

wp_year *Function to extract the year*

Description

Function to extract the year from a timestamp of e.g. class POSIXlt, POSIXct, Date, or character. If the option is of type character it should be in the form of yyyy-mm-dd.

Usage

```
wp_year(timestamp)

## S3 method for class 'POSIXlt'
wp_year(timestamp)

## Default S3 method:
wp_year(timestamp)
```

Arguments

timestamp Some sort of timestamp e.g. of class POSIXlt, POSIXct, Date, or character. If the option is of type character it should be in the form of yyyy-mm-dd.

Methods (by class)

- POSIXlt: extract
- default: extract

wp_yearmonth *function for getting year and month of a timestamp*

Description

function for getting year and month of a timestamp

Usage

wp_yearmonth(timestamp)

Arguments

timestamp Some sort of timestamp e.g. of class POSIXlt, POSIXct, Date, or character. If the option is of type character it should be in the form of yyyy-mm-dd.

write_utf8_csv *function to write csv files with UTF-8 characters (even under Windwos)*

Description

function to write csv files with UTF-8 characters (even under Windwos)

Usage

write_utf8_csv(df, file)

Arguments

df data frame to be written to file
file file name / path where to put the data

Index

*Topic **datasets**

wp_cache, 5

html, 2

html2, 3

prev_month_end, 3

prev_month_start, 3

print.wp_df, 4

read_utf8_csv, 4

toUTF8, 4

wp_add_to_cache, 5

wp_cache, 5

wp_cache_default, 5

wp_cache_file, 6

wp_cache_load, 6

wp_cache_reset, 6

wp_check_date_inputs, 7

wp_date, 7

wp_day, 8

wp_download_data, 9

wp_expand_ts, 10

wp_get_cache, 10

wp_get_data, 10

wp_http_header, 11

wp_jsons_to_df, 11

wp_linked_pages, 11

wp_load, 12

wp_month, 12

wp_prepare_urls, 13

wp_save, 13

wp_save_cache, 14

wp_set_cache_file, 14

wp_set_print_options, 14

wp_trend, 15

wp_wday, 16

wp_year, 16

wp_yearmonth, 17

write_utf8_csv, 17