Package ‘wikipediatrend’

June 3, 2020

Title Public Subject Attention via Wikipedia Page View Statistics
Version 2.1.6
Date 2020-06-03
Maintainer Peter Meissner <retep.meissner@gmail.com>
Description A convenience wrapper for the Wikipedia page access statistics API
    binding the 'pageviews' package and using an additional self composed data
    source thus covering a time span from very late 2007 up to the present for
daily page views.
License GPL (>= 2)
Depends R (>= 3.0.0)
Imports pageviews (>= 0.3.0), stringr (>= 0.6.2), rvest (>= 0.2.0),
    httr (>= 0.6.1), utils, xml2 (>= 0.1.2), hellno (>= 0.0.1),
    glue (>= 1.1.1), ggplot2 (>= 2.2.1)
Suggests testthat, knitr
BugReports https://github.com/petermeissner/wikipediatrend/issues
RoxxygenNote 7.0.2
NeedsCompilation no
Author Peter Meissner [aut, cre],
    R Core Team [cph]
Repository CRAN
Date/Publication 2020-06-03 15:10:02 UTC

R topics documented:

plot.wp_df ................................................................. 2
print.wp_df .............................................................. 2
wiki_search ............................................................... 3
wpd_decompress ......................................................... 3
wpd_get_exact ........................................................... 4
wpd_search ............................................................... 4
wp_check_date_inputs ................................................... 5
Description

simple plot function for objects of class wp_df

Usage

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

Arguments

- `x`: object of class wp_df to be plotted
- `...`: other parameter

Description

a wikipediastrend specific data frame

Usage

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

Arguments

- `x`: the thing to be printed
- `...`: print.default parameters
**Description**

wiki_search

**Usage**

```py
wiki_search(page_fragment = "R_(programming_language)", lang = "en")
```

**Arguments**

- `page_fragment`: substring to search for
- `lang`: wikipedia language to search in

---

**Description**

wpd_decompress

**Usage**

```py
wpd_decompress(cont_item, lang)
```

**Arguments**

- `cont_item`: single content item
- `lang`: language to decompress
### wpd_get_exact

**Description**

wpd_get_exact

**Usage**

wpd_get_exact(page, lang, from, to, warn = TRUE)

**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), then the page name equals to Peter_Fox_(musician).
- **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/.
- **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.
- **warn**: Warn about not being able to retrieve data or not?

### wpd_search

**Description**

wpd_search

**Usage**

wpd_search(page_fragment = "R_(programming_language)", lang = "en")

**Arguments**

- **page_fragment**: substring to search for within titles
- **lang**: language to search
**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_check_page_input**

*check page input*

**Description**

check page input

**Usage**

wp_check_page_input(page)

**Arguments**

- **page**: the pages 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
...
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()`

---

**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**

```r
wp_day(timestamp)
```

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

```r
## 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_get_data  

*function for getting data (download + extraction)*

**Description**

function for getting data (download + extraction)

**Usage**

```r
wp_get_data(
  page = "R_(programming_language)",
  from = "2007-12-01",
  to = as.character(Sys.Date()),
  lang = "en",
  user_type = "all",
  platform = "all",
  warn = TRUE
)
```

**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), then 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/.
- **user_type**: the type of users. One or more of "all", "user", "spider" or "bot". "all" by default.
- **platform**: The platform the pageviews came from: One or more of "all", "desktop", "mobile-web" and "mobile-app". Set to "all" by default.
- **warn**: Warn about not being able to retrieve data or not?
wp_linked_pages

function looking for other languages of a page

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 given title

wp_month

Function to extract the month

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
Description

Function for getting access statistics for wikipedia pages

Usage

wp_trend(
    page = "R_(programming_language)",
    from = "2007-12-10",
    to = Sys.Date(),
    lang = "en",
    warn = TRUE
)

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/.

warn Warn about not being able to retrieve data or not?

Examples

res <- wp_trend(page=c("Der_Spiegel", "Die_Zeit"), lang=c("de", "en"))
**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**

```r
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 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**

```r
wp_year(timestamp)
## S3 method for class 'POSIXlt'
wp_year(timestamp)
## Default S3 method:
wp_year(timestamp)
```
wp_yearmonth

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.
Index

plot.wp_df, 2
print.wp_df, 2

wiki_search, 3
wp_check_date_inputs, 5
wp_check_page_input, 5
wp_date, 6
wp_day, 7
wp_get_data, 8
wp_linked_pages, 9
wp_month, 9
wp_trend, 10
wp_wday, 11
wp_year, 11
wp_yearmonth, 12
wpd_decompress, 3
wpd_get_exact, 4
wpd_search, 4