Package ‘wikipediatrend’

October 12, 2022

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
RoxygenNote 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
plot.wp_df

simple plot function for objects of class wp_df

Description

simple plot function for objects of class wp_df

Usage

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

Arguments

x  object of class wp_df to be plotted
...
other parameter

print.wp_df

a wikipediastrend specific data frame

Description

a wikipediastrend specific data frame

Usage

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

Arguments

x  the thing to be printed
...
print.default parameters
**wiki_search**

**Description**

wiki_search

**Usage**

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

**Arguments**

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

---

**wpd_decompress**

**Description**

wpd_decompress

**Usage**

```python
wpd_decompress(cont_item, lang)
```

**Arguments**

- `cont_item`: single content item
- `lang`: language to decompress
Description

wpd_get_exact

Usage

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

Arguments

<table>
<thead>
<tr>
<th>Argument</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>page</td>
<td>The name of the Wikipedia page as to be found in the URL to the article. If e.g. the URL is: <a href="https://en.wikipedia.org/wiki/Peter_Fox_(musician)">https://en.wikipedia.org/wiki/Peter_Fox_(musician)</a>, than the page name equals to Peter_Fox_(musician).</td>
</tr>
<tr>
<td>lang</td>
<td>The language shorthand identifying which Wikipedia access statistics are to be used: e.g. &quot;en&quot; for the English version found at <a href="https://en.wikipedia.org">https://en.wikipedia.org</a>, &quot;de&quot; for the German version found at <a href="https://de.wikipedia.org">https://de.wikipedia.org</a> or perhaps &quot;als&quot; for the Alemannic dialect found under <a href="https://als.wikipedia.org/">https://als.wikipedia.org/</a>.</td>
</tr>
<tr>
<td>from</td>
<td>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.</td>
</tr>
<tr>
<td>to</td>
<td>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.</td>
</tr>
<tr>
<td>warn</td>
<td>Warn about not being able to retrieve data or not?</td>
</tr>
</tbody>
</table>

Description

wpd_search

Usage

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

Arguments

<table>
<thead>
<tr>
<th>Argument</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>page_fragment</td>
<td>substring to search for within titles</td>
</tr>
<tr>
<td>lang</td>
<td>language to search</td>
</tr>
</tbody>
</table>
**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**

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

**Arguments**

- **from**: first date of timespan to check
- **to**: second date of timespan to check

---

**Description**

check page input

**Usage**

```python
wp_check_page_input(page)
```

**Arguments**

- **page**: the pages to check
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()

---

**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**

```r
wp_day(timestamp)
```

### ## S3 method for class 'POSIXlt'

```r
wp_day(timestamp)
```

### ## Default S3 method:

```r
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

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), 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/.

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?
### 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

### 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)

```r
## 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_trend

Function for getting access statistics for wikipedia pages

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
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
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

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