

Package ‘newsanchor’

June 29, 2019

Type Package

Title Client for the News API

Version 0.1.1

Description Interface to gather news from the 'News API', based on a multi-level query <<https://newsapi.org/>>. A personal API key is required.

License MIT + file LICENSE

Encoding UTF-8

Depends R (>= 3.2.0)

BugReports <https://github.com/correlaid/newsanchor/issues>

LazyData true

Imports devtools, httr, jsonlite, tidyr, xml2, lubridate, askpass

Suggests dplyr, knitr, magrittr, rmarkdown, robotstxt, rvest, stringr, testthat, mockery, tidytext, textdata

RoxygenNote 6.1.1

VignetteBuilder knitr

NeedsCompilation no

Author Preu Frie [aut, pro],
Buhl Yannik [aut, cre],
Schulze Lars [aut],
Dix Jan [aut, pro]

Maintainer Buhl Yannik <ybuhl@posteo.de>

Repository CRAN

Date/Publication 2019-06-29 13:10:03 UTC

R topics documented:

build_newsanchor_url	2
collapse_to_comma_separated	3
extract_newsanchor_articles	3
extract_newsanchor_metadata	4

extract_newsanchor_sources	4
get_everything	5
get_everything_all	6
get_headlines	8
get_headlines_all	9
get_sources	10
make_newsanchor_get_request	11
parse_newsanchor_content	11
sample_response	12
set_api_key	13
stop_if_invalid_category	13
stop_if_invalid_country	14
stop_if_invalid_language	14
stop_if_invalid_source	15
terms_category	15
terms_country	16
terms_language	16
terms_sources	16
Index	17

build_newsanchor_url *Builds query URL for newsapi.org.*

Description

build_newsanchor_url adds a list of query arguments to a given News API endpoint.

Usage

```
build_newsanchor_url(url, query_args)
```

Arguments

url	NEWS API endpoint.
query_args	named list of parameters that are needed to query the endpoint. Check the News API documentation to see which endpoint requires which parameters.

Value

httr URL.

`collapse_to_comma_separated`*Concatenate character vector to comma-separated string.*

Description

`collapse_to_comma_separated` is a helper function that concatenates a character vector to a comma-separated string. If the input vector has only one element, the element will be returned unchanged.

Usage

```
collapse_to_comma_separated(v)
```

Arguments

`v` character vector.

Value

string with elements of `v` separated by comma.

`extract_newsanchor_articles`*Extracts data frame with News API articles from response object.*

Description

`extract_newsanchor_articles` extracts a data frame containing the News API articles that matched the request to News API everything or headlines endpoint.

Usage

```
extract_newsanchor_articles(metadata, content_parsed)
```

Arguments

`metadata` data frame containing meta data related to the request, see `extract_newsanchor_metadata`.
`content_parsed` parsed content of a response to News API query

Value

data frame containing articles.

extract_newsanchor_metadata

Extracts metadata.

Description

extract_newsanchor_metadata extracts meta data from the response object and the parsed content.

Usage

```
extract_newsanchor_metadata(response, content_parsed, page = NULL,
                             page_size = NULL)
```

Arguments

response	httr response object
content_parsed	parsed content of a response to News API query
page	Specifies the page number of your results that was returned. Defaults to NULL.
page_size	The number of articles per page that were returned. Defaults to NULL.

Value

data frame containing meta data related to the query.

extract_newsanchor_sources

Extracts data frame with News API sources from response object.

Description

extract_newsanchor_sources extracts a data frame containing the News API sources that matched the request to News API sources endpoint.

Usage

```
extract_newsanchor_sources(metadata, content_parsed)
```

Arguments

metadata	data frame containing meta data related to the request, see extract_newsanchor_metadata.
content_parsed	parsed content of a response to News API query

Value

data frame containing sources.

get_everything *Get resources of newsapi.org*

Description

get_everything returns articles from large and small news sources and blogs. This includes news as well as other regular articles. You can search for multiple sources, different language, or use your own keywords. Articles can be sorted by the earliest date publishedAt, relevancy, or popularity. To automatically download all results, use get_everything_all().

Please check that the api_key is available. You can provide an explicit definition of the key or use set_api_key().

Valid languages for language are provided in the dataset terms_language.

Usage

```
get_everything(query, sources = NULL, domains = NULL,
  exclude_domains = NULL, from = NULL, to = NULL, language = NULL,
  sort_by = "publishedAt", page = 1, page_size = 100,
  api_key = Sys.getenv("NEWS_API_KEY"))
```

Arguments

query	Character string that contains the searchterm for the API's data base. API supports advanced search parameters, see 'details'. Passing a searchterm is compulsory.
sources	Character vector with with IDs of the news outlets you want to focus on (e.g., c("usa-today", "spiegel-online")).
domains	Character vector with domains that you want to restrict your search to (e.g. c("bbc.com", "nytimes.com")).
exclude_domains	Similar usage as with 'domains'. Will exclude these domains from your search.
from	Character string with start date of your search. Needs to conform to one of the following lubridate order strings: "ymdHMs, ymdHMsZ, ymd". See help for lubridate::parse_date_time. If from is not specified, NewsAPI defaults to the oldest available date (depends on your paid/unpaid plan from newsapi.org).
to	Character string that marks the end date of your search. Needs to conform to one of the following lubridate order strings: "ymdHMs, ymdHMsZ, ymd". See help for lubridate::parse_date_time. If to is not specified, NewsAPI defaults to the most recent article available.
language	Specifies the language of the articles of your search. Must be in ISO shortcut format (e.g., "de", "en"). See list of all languages using newsanchor::terms_language. Default is all languages.

sort_by	Character string that specifies the sorting variable of your article results. Accepts three options: "publishedAt", "relevancy", "popularity". Default is "publishedAt".
page	Specifies the page number of your results that is returned. Must be numeric. Default is first page. If you want to get all results at once, use get_everything_all from 'newsanchor'.
page_size	The number of articles per page that are returned. Maximum is 100 (also default).
api_key	Character string with the API key you get from newsapi.org. Passing it is compulsory. Alternatively, function can be provided from the global environment (see set_api_key()).

Details

Advanced search (see also www.newsapi.org): Surround entire phrases with quotes (") for exact matches. Prepend words/phrases that must appear with "+" symbol (e.g., +bitcoin). Prepend words that must not appear with "-" symbol (e.g., -bitcoin). You can also use AND, OR, NOT keywords (optionally grouped with parenthesis, e.g., 'crypto AND (ethereum OR litecoin) NOT bitcoin').

Value

List with two dataframes:

- 1) Data frame with results_df
- 2) Data frame with meta_data

Examples

```
## Not run:
df <- get_everything(query = "stuttgart", language = "de")
df <- get_everything(query = "mannheim", from = "2019-01-02 12:00:00")

## End(Not run)
```

get_everything_all *Returns all articles from newsapi.org in one data frame*

Description

get_everything searches through articles from large and small news sources and blogs. This includes breaking news as well as other regular articles. You can search for multiple sources, different language, or use your own keywords. Articles can be sorted by the earliest date publishedAt, relevancy, or popularity. To automatically download all results, use get_everything_all()

Please check that the api_key is available. You can provide an explicit definition of the api_key or use set_api_key().

Valid languages for language are provided in the dataset

terms_language. To automatically download all results for one search, use get_everything_all

. Please check that the api_key is available. You can provide an explicit definition of the api_key or use set_api_key

For valid searchterms see data(searchterms)

Usage

```
get_everything_all(query, sources = NULL, domains = NULL,
  exclude_domains = NULL, from = NULL, to = NULL, language = NULL,
  sort_by = "publishedAt", api_key = Sys.getenv("NEWS_API_KEY"))
```

Arguments

query	Character string that contains the searchterm for the API's data base. API supports advanced search parameters, see 'details'.
sources	Character string with IDs (comma separated) of the news outlets you want to focus on (e.g., "usa-today, spiegel-online").
domains	Character string (comma separated) with domains that you want to restrict your search to (e.g., "bbc.com, nytimes.com").
exclude_domains	Similar usage as with 'domains'. Will exclude these domains from your search.
from	Marks the start date of your search. Must be in ISO 8601 format (e.g., "2018-09-08" or "2018-09-08T12:51:42"). Default is the oldest available date (depends on your paid/unpaid plan from newsapi.org).
to	Marks the end date of your search. Works similarly to 'from'. Default is the latest article available.
language	Specifies the language of the articles of your search. Must be in ISO shortcut format (e.g., "de", "en"). See list of all languages on https://newsapi.org/docs/endpoints/everything . Default is all languages.
sort_by	Character string that specifies the sorting of your article results. Accepts three options: "publishedAt", "relevancy", "popularity". Default is "publishedAt".
api_key	Character string with the API key you get from newsapi.org. Passing it is compulsory. Alternatively, function can be provided from the global environment (see set_api_key).

Value

List with two dataframes:

- 1) Data frame with results_df
- 2) Data frame with meta_data

Examples

```
## Not run:
df <- get_everything_all(query = "mannheim")
df <- get_everything_all(query = "stuttgart", language = "en")

## End(Not run)
```

get_headlines

Returns selected headlines from newsapi.org

Description

get_headlines returns live top and breaking headlines for a country, specific category in a country, single source, or multiple sources. You can also search with keywords. Articles are sorted by the earliest date published first. To automatically download all results, use get_headlines_all().

Please check that the api_key is available. You can provide an explicit definition of the key or use set_api_key().

Valid searchterms are provided in the data sets terms_category, terms_country or terms_sources.

Usage

```
get_headlines(query = NULL, category = NULL, country = NULL,
  sources = NULL, page = 1, page_size = 100,
  api_key = Sys.getenv("NEWS_API_KEY"))
```

Arguments

query	Character string that contains the searchterm.
category	Character string with the category you want headlines from.
country	Character string with the country you want headlines from.
sources	Character vector with with IDs of the news outlets you want to focus on (e.g., c("usa-today", "spiegel-online")).
page	Specifies the page number of your results that is returned. Must be numeric. Default is first page. If you want to get all results at once, use get_headlines_all from 'newsanchor'.
page_size	The number of articles per page that are returned. Maximum is 100 (also default).
api_key	Character string with the API key you get from newsapi.org. Passing it is compulsory. Alternatively, a function can be provided from the global environment (see set_api_key).

Value

List with two dataframes:

- 1) Data frame with results_df
- 2) Data frame with meta_data

Examples

```
## Not run:  
df <- get_headlines(sources = "bbc-news")  
df <- get_headlines(query = "sports", page = 2)  
df <- get_headlines(category = "business")  
  
## End(Not run)
```

get_headlines_all *Returns all headlines from newsapi.org*

Description

get_headlines returns live top and breaking headlines for a country, specific category in a country, single source, or multiple sources. You can also search with keywords. Articles are sorted by the earliest date published first. To automatically download all results, use get_headlines_all.

Please check that the api_key is available. You can provide an explicit definition of the api_key or use set_api_key

Valid searchterms are provided in terms_category, terms_country or terms_sources

Usage

```
get_headlines_all(query = NULL, category = NULL, country = NULL,  
  sources = NULL, api_key = Sys.getenv("NEWS_API_KEY"))
```

Arguments

query	Character string that contains the searchterm
category	Category you want headlines from
country	Country you want headlines for
sources	Character string with IDs (comma separated) of the news outlets you want to focus on (e.g., "usa-today, spiegel-online").
api_key	Character string with the API key you get from newsapi.org. Passing it is compulsory. Alternatively, function can be provided from the global environment (see set_api_key).

Value

List with two dataframes:

- 1) Data frame with results_df
- 2) Data frame with meta_data

Examples

```
## Not run:  
df <- get_headlines_all(query = "sports")  
df <- get_headlines_all(category = "health")  
  
## End(Not run)
```

get_sources

Returns selected sources from newsapi.org

Description

get_sources returns the news sources currently available on newsapi.org. The sources can be filtered using category, language or country. If the arguments are empty the query return all available sources.

Usage

```
get_sources(category = NULL, language = NULL, country = NULL,  
            api_key = Sys.getenv("NEWS_API_KEY"))
```

Arguments

category	Category you want to get sources for as a string. Default: NULL.
language	The language you want to get sources for as a string. Default: NULL.
country	The country you want to get sources for as a string (e.g. "us"). Default: NULL.
api_key	String with the API key you get from newsapi.org. Passing it is compulsory. Alternatively, function can be provided from the global environment (see set_api_key).

Value

List with two dataframes:

- 1) Data frame with results_df
- 2) Data frame with meta_data

Examples

```
## Not run:
get_sources(api_key)
get_sources(api_key, category = "technology")
get_sources(api_key, language = "en")

## End(Not run)
```

make_newsanchor_get_request

Makes a GET request to News API.

Description

make_newsanchor_get_request makes a GET request to News API.

Usage

```
make_newsanchor_get_request(url, api_key)
```

Arguments

url News API url with query parameters and scheme specified. See build_newsanchor_url.
api_key News API key.

Value

httr response object.

parse_newsanchor_content

Parses content returned by query to the News API.

Description

parse_newsanchor_content parses the content sent back by the News API to an R list.

Usage

```
parse_newsanchor_content(response)
```

Arguments

response httr response object

Value

R list.

sample_response	<i>Sample Response Object</i>
-----------------	-------------------------------

Description

A sample response object generated using 'get_everything'.

Usage

```
sample_response
```

Format

An object of class list of length 2.

Details

This response object was mainly created for demonstrating purposes. The data set is used in the "Scrape New York Times Online Articles" vignette. The object was created using the following query.

Value

List with two dataframes:
1) Data frame with results_df
2) Data frame with meta_data

Examples

```
## Not run:  
response <- get_everything(query = "Trump",  
                           sources = "the-new-york-times",  
                           from = "2018-12-03",  
                           to = "2018-12-09")  
  
## End(Not run)
```

set_api_key	<i>Add API key to the .Renviron</i>
-------------	-------------------------------------

Description

Function to set you API Key to the R environment when starting using newsanchor package. Attention: You should only execute this functions once.

Usage

```
set_api_key(path = stop("Please specify a path."))
```

Arguments

path character. Path where the environment is stored. Default is the normalized path.

Value

None.

Author(s)

Jan Dix <<jan.d@correlaid.org>>

Examples

```
## Not run:
set_api_key(tempdir()) # you will be prompted to enter your API key.

## End(Not run)
```

stop_if_invalid_category	<i>Checks validity of a category.</i>
--------------------------	---------------------------------------

Description

stop_if_invalid_category checks whether a given category is valid for News API and stops with an error if this is not the case.

Usage

```
stop_if_invalid_category(category)
```

Arguments

category category to check as a string.

`stop_if_invalid_country`*Checks validity of a country*

Description

`stop_if_invalid_country` checks whether a given country is valid for News API and stops with an error if this is not the case.

Usage

```
stop_if_invalid_country(country)
```

Arguments

`country` country to check as a string.

`stop_if_invalid_language`*Checks validity of a language*

Description

`stop_if_invalid_language` checks whether a given language is valid for News API and stops with an error if this is not the case.

Usage

```
stop_if_invalid_language(language)
```

Arguments

`language` language to check as a string.

`stop_if_invalid_source`*Checks validity of a source*

Description

`stop_if_invalid_source` checks whether a given source is valid for News API and stops with an error if this is not the case.

Usage

```
stop_if_invalid_source(source)
```

Arguments

`source` source to check as a string.

`terms_category`*Terms Category*

Description

The dataframe provides possible categories (e.g., sports) you want to get headlines for. This dataframe is relevant in conjunction with `get_headlines`.

Usage

```
terms_category
```

Format

An object of class `data.frame` with 7 rows and 1 columns.

terms_country	<i>Terms Country</i>
---------------	----------------------

Description

This dataframe provides possible countries you want to get news from. This dataframe is relevant in conjunction with `get_headlines`.

Usage

```
terms_country
```

Format

An object of class `data.frame` with 54 rows and 1 columns.

terms_language	<i>Terms Language</i>
----------------	-----------------------

Description

This dataframe provides possible languages you want to get news for. This dataframe is relevant in conjunction with `get_everything`.

Usage

```
terms_language
```

Format

An object of class `data.frame` with 14 rows and 1 columns.

terms_sources	<i>Terms Sources</i>
---------------	----------------------

Description

This dataframe provides possible news sources or blogs you want to get news from. This dataframe is relevant in conjunction with `get_everything`.

Usage

```
terms_sources
```

Format

An object of class `data.frame` with 138 rows and 1 columns.

Index

*Topic **datasets**

sample_response, [12](#)

terms_category, [15](#)

terms_country, [16](#)

terms_language, [16](#)

terms_sources, [16](#)

build_newsanchor_url, [2](#)

collapse_to_comma_separated, [3](#)

extract_newsanchor_articles, [3](#)

extract_newsanchor_metadata, [4](#)

extract_newsanchor_sources, [4](#)

get_everything, [5](#)

get_everything_all, [6](#)

get_headlines, [8](#)

get_headlines_all, [9](#)

get_sources, [10](#)

make_newsanchor_get_request, [11](#)

parse_newsanchor_content, [11](#)

sample_response, [12](#)

set_api_key, [13](#)

stop_if_invalid_category, [13](#)

stop_if_invalid_country, [14](#)

stop_if_invalid_language, [14](#)

stop_if_invalid_source, [15](#)

terms_category, [15](#)

terms_country, [16](#)

terms_language, [16](#)

terms_sources, [16](#)