

Package ‘newspapr’

November 26, 2019

Title API Wrapper for NewsAPI

Version 0.0.0.9000

Description A set of functions for using NewsAPI from within R.

BugReports <https://github.com/mcgirjau/newspapr/issues>

Depends R (≥ 3.1.0)

License GPL-2

Encoding UTF-8

LazyData true

RoxygenNote 7.0.0

Imports curl,
dplyr,
httr,
janitor,
parsedate,
stringr,
tibble

Suggests testthat

R topics documented:

| | |
|-----------------------------|----------|
| check_key | 2 |
| get_everything | 2 |
| get_sources | 3 |
| get_top_headlines | 4 |
| has_key | 5 |
| register_key | 6 |
| show_key | 6 |
| Index | 7 |

| | |
|-----------|---------------------------------------|
| check_key | <i>Check whether API key is valid</i> |
|-----------|---------------------------------------|

Description

Checks whether an API key is valid. Prints a message to let the user know whether their key is usable with NewsAPI (case in which the function prints "All OK"), or whether an error occurred (case in which the function prints a descriptive message of said error).

Usage

```
check_key(key = NULL)
```

Arguments

key (Optional) A string. Your NewsAPI key.

| | |
|----------------|-------------------------|
| get_everything | <i>Get all articles</i> |
|----------------|-------------------------|

Description

Search through millions of articles from over 30,000 large and small news sources and blogs. This includes breaking news as well as lesser articles. At least one parameter from among keyword, title_keyword, sources, and domains must be specified, otherwise the search is too broad.

Usage

```
get_everything(
  keyword = NULL,
  title_keyword = NULL,
  sources = NULL,
  domains = NULL,
  exclude_domains = NULL,
  from = NULL,
  to = NULL,
  language = NULL,
  sort_by = "published at",
  page_size = NULL,
  page = NULL
)
```

Arguments

keyword (Optional) A string. Keyword(s) based on which to search for articles, such as "Hong Kong protests" or "Amazon forest". Keywords may appear either in the title or the body of the article. Advanced search is supported here - surround phrases with quotes (") for exact match, prepend words or phrases that must appear with a + symbol (e.g. +bitcoin), prepend words

| | |
|-----------------|---|
| | that must not appear with a - symbol (e.g. -bitcoin). Alternatively, you can use the AND / OR / NOT keywords, and optionally group these with parenthesis. Eg: "crypto AND (ethereum OR litecoin) NOT bitcoin". |
| title_keyword | (Optional) A string or character vector. Keywords or phrases to search for in the article title only. Advanced search is supported here (see parameter "keyword") for details. |
| sources | (Optional) A character vector. The news sources or blogs you want top headlines from. See the get_sources() function for how to obtain a full list of the available sources. |
| domains | (Optional) A character vector. The set of domains (eg bbc.co.uk, techcrunch.com, engadget.com) to restrict the search to. |
| exclude_domains | (Optional) A character vector. The set of domains (eg bbc.co.uk, techcrunch.com, engadget.com) to remove from the results. |
| from | (Optional) A string. A date and optional time for the oldest article allowed. This should be in ISO 8601 format (e.g. "2019-11-25" or "2019-11-25T23:45:05"). Default: the oldest according to your plan. |
| to | (Optional) A string. A date and optional time for the newest article allowed. This should be in ISO 8601 format (e.g. "2019-11-25" or "2019-11-25T23:45:05"). Default: the newest according to your plan. |
| language | (Optional) A string. Find sources that display news in a specific language. Possible options are: Arabic, German, English, Spanish, French, Hebrew, Italian, Dutch, Norwegian, Portuguese, Russian, Swedish, Urdu, and Chinese. The language can also be specified using ISO codes, e.g. "DE" for German or "ES" for Spanish. By default, the API returns sources in all languages. |
| sort_by | (Optional) A string. The order to sort the articles in. Possible options: "relevancy", "popularity", "published at". Defaults to "published at". |
| page_size | (Optional) An integer. The number of results to return per request. The API's default is 20, and the maximum is 100. |
| page | (Optional) An integer. Use this to page through the results if the total results found is greater than the page size. |

Value

A tibble of article information, with source, author, title, description, URL, publication date, and a short preview of the content.

| | |
|-------------|---------------------------------------|
| get_sources | <i>Get index of available sources</i> |
|-------------|---------------------------------------|

Description

Returns the subset of news publishers that top headlines are available from. It's mainly a convenience function that can be used to keep track of the publishers available on the API. It seems a large amount of sources that do return top headlines are not included by NewsAPI in their output. As such, these sources cannot be used in a query, although they do return results. The complete index of sources that can be used in a NewsAPI query is obtained using get_sources().

Usage

```
get_sources(category = NULL, language = NULL, country = NULL)
```

Arguments

| | |
|----------|--|
| category | (Optional) A string. Find sources that display news of a specific category. Possible options are: "business", "entertainment", "general", "health", "science", "sports", and "technology". By default, the API returns sources of all categories. |
| language | (Optional) A string. Find sources that display news in a specific language. Possible options are: Arabic, German, English, Spanish, French, Hebrew, Italian Dutch, Norwegian, Portuguese, Russian, Swedish, Urdu, and Chinese. The language can also be specified using ISO codes, e.g. "DE" for German or "ES" for Spanish. By default, the API returns sources in all languages. |
| country | (Optional) A string. The 2-letter ISO 3166-1 code of the country you want to get headlines for, such as "US" or "FR". This parameter cannot be mixed with the sources parameter. |

Value

A tibble of news publishers, with their ID as it should be queried in a NewsAPI request, a short description, their URL, their category, their language, and their country.

Examples

```
get_sources(language = "German")
```

| | |
|--------------------------------|-----------------------------------|
| <code>get_top_headlines</code> | <i>Get top headlines articles</i> |
|--------------------------------|-----------------------------------|

Description

Returns a list of current top headlines articles based on some query. At least one parameter from among keyword, country, category, and sources must be specified, with the limitation that the sources parameter cannot be mixed with country or category.

Usage

```
get_top_headlines(
  keyword = NULL,
  country = NULL,
  category = NULL,
  sources = NULL,
  page_size = NULL,
  page = NULL
)
```

Arguments

| | |
|-----------|---|
| keyword | (Optional) A string. Keyword(s) based on which to search for articles, such as "Hong Kong protests" or "Amazon forest". |
| country | (Optional) A string. The 2-letter ISO 3166-1 code of the country you want to get headlines for, such as "US" or "FR". This parameter cannot be mixed with the sources parameter. |
| category | (Optional) A string. The category you want to get top headlines for, such as "business", "entertainment", or "sports". This parameter cannot be mixed with the sources parameter. |
| sources | (Optional) A character vector. The news sources or blogs you want top headlines from. See the <code>get_sources()</code> function for how to obtain a full list of the available sources. This parameter cannot be mixed with the <code>country_iso</code> or <code>category</code> parameters. |
| page_size | (Optional) An integer. The number of results to return per request. The API's default is 20, and the maximum is 100. |
| page | (Optional) An integer. Use this to page through the results if the total results found is greater than the page size. |

Value

A tibble of article information, with source, author, title, description, URL, publication date, and a short preview of the content.

Examples

```
get_top_headlines(keyword = "Panama Canal", page_size = 10, page = 2)
get_top_headlines(country = "br", category = "sports")
get_top_headlines(keyword = "iPhone 10", source = "techcrunch")
```

has_key

Check whether an API key has been registered

Description

Checks whether an API key has been registered within the current R session, either manually by using `register_key()`, or automatically from the `.Renv` file.

Usage

```
has_key()
```

Value

A boolean indicating whether an API key has been registered within the current R session.

| | |
|--------------|-------------------------|
| register_key | <i>Register API key</i> |
|--------------|-------------------------|

Description

Returns the subset of news publishers that top headlines are available from. It's mainly a convenience function that can be used to keep track of the publishers available on the API.

Usage

```
register_key(key, write = FALSE)
```

Arguments

| | |
|-------|---|
| key | A string. Your NewsAPI key. |
| write | (Optional) A boolean. Specifies whether you want your API key to be written to your .Renviron file for use in all future R sessions. Defaults to FALSE. |

| | |
|----------|------------------------|
| show_key | <i>Display API key</i> |
|----------|------------------------|

Description

Displays the API key that has been registered.

Usage

```
show_key()
```

Index

check_key, [2](#)

get_everything, [2](#)

get_sources, [3](#)

get_top_headlines, [4](#)

has_key, [5](#)

register_key, [6](#)

show_key, [6](#)