

Package ‘rzeit2’

January 7, 2019

Type Package

Title Client for the ZEIT ONLINE Content API

Version 0.2.3

Description

Interface to gather newspaper articles from 'DIE ZEIT' and 'ZEIT ONLINE', based on a multi-level query <<http://developer.zeit.de/>>. A personal API key is required for usage.

License MIT + file LICENSE

Encoding UTF-8

Depends R (>= 3.2.0)

BugReports <https://github.com/jandix/rzeit2/issues>

LazyData true

RoxygenNote 6.1.1

Imports anytime, httr, jsonlite, openssl, rvest, stringr, xml2

Suggests dplyr, ggplot2, ggthemes, knitr, rmarkdown, robotstxt, tidytext

VignetteBuilder knitr

NeedsCompilation no

Author Jan Dix [aut, cre]

Maintainer Jan Dix <jan.dix@uni-konstanz.de>

Repository CRAN

Date/Publication 2019-01-07 18:00:03 UTC

R topics documented:

get_article_comments	2
get_article_images	3
get_article_text	4
get_client	5
get_content	5
get_content_all	7

rzeit2	8
sentiment_example	8
senti_ws	9
set_api_key	9

Index	11
--------------	-----------

get_article_comments *Get article comments*

Description

Get the article comments for a single url.

Usage

```
get_article_comments(url, id = NULL, simplify = FALSE,
  timeout = NULL)
```

Arguments

url	character. A single character string or character vector.
id	character. You can provide your own id for each article. If is null the function uses the md5 hash of the url to create one.
simplify	logical. If true the function returns a data frame else it returns a nested list.
timeout	integer. Seconds to wait between queries.

Details

get_article_comments is the function, which fetches and parses article comments. This function may break in the future due to layout changes on the ZEIT ONLINE website.

Value

A list with comments and their respective replies. If the content lies beyond the paywall the function returns "[ZEIT PLUS CONTENT] You need a ZEIT PLUS account to access this content."

Warning

Please use that function carefully because it uses a lot of HTTP requests. The extensive usage of this function may result in the blocking of IP.

Author(s)

Jan Dix <<jan.dix@uni-konstanz.de>>

Examples

```
## Not run:
url <- paste0("https://www.zeit.de/kultur/film/2018-04/",
"tatort-frankfurt-unter-kriegern-obduktionsbericht")
get_article_comments(url = url)

## End(Not run)
```

get_article_images *Get article images*

Description

Get the article images for a single url or a vector of urls.

Usage

```
get_article_images(url, timeout = 0, download = NULL)
```

Arguments

url	character. A single character string or character vector.
timeout	integer. Seconds to wait between queries.
download	character. Path to download folder. If path is set to NULL images are not downloaded.

Details

get_article_images is the function, which fetches and parses meta information for each image of an article and downloads the images. This function may break in the future due to layout changes on the ZEIT ONLINE website.

Value

A data frame including meta information for each image.

Author(s)

Jan Dix <<jan.dix@uni-konstanz.de>>

Examples

```
## Not run:
url <- paste0("https://www.zeit.de/kultur/film/2018-04/",
"tatort-frankfurt-unter-kriegern-obduktionsbericht")
get_article_images(url = url, timeout = 0)

## End(Not run)
```

get_article_text	<i>Get article text</i>
------------------	-------------------------

Description

Get the article text for a single url or a vector of urls.

Usage

```
get_article_text(url, timeout = NULL)
```

Arguments

url	character. A single character string or character vector.
timeout	integer. Seconds to wait between queries.

Details

get_article_text is the function, which fetches and parses an article. This function may break in the future due to layout changes on the ZEIT ONLINE website.

Value

A named character vector with the respective text. If the content lies beyond the paywall the function returns "[ZEIT PLUS CONTENT] You need a ZEIT PLUS account to access this content."

Author(s)

Jan Dix <<jan.dix@uni-konstanz.de>>

Examples

```
## Not run:
url <- paste0("https://www.zeit.de/kultur/film/2018-04/",
"tatort-frankfurt-unter-kriegern-obduktionsbericht")
get_article_text(url = url)

## End(Not run)
```

get_client	<i>Client status and API usage</i>
------------	------------------------------------

Description

get_client returns API access status and usage.

Usage

```
get_client(api_key = Sys.getenv("ZEIT_ONLINE_KEY"))
```

Arguments

api_key character. The personal api code. To request an API key see: <http://developer.zeit.de/quickstart/> This parameter is by default set to the R Environment.

Value

a list of information about the client and API usage

Author(s)

Jan Dix (<jan.dix@uni-konstanz.de>)

Examples

```
## Not run:  
get_client()  
  
## End(Not run)
```

get_content	<i>Content endpoint</i>
-------------	-------------------------

Description

Exposes a search in the ZEIT online archive on the content endpoint and returns results for the given query.

Usage

```
get_content(query, limit = 10, offset = 0, sort = "release_date asc",  
begin_date = NULL, end_date = NULL,  
api_key = Sys.getenv("ZEIT_ONLINE_KEY"))
```

Arguments

query	character. Search query term.
limit	integer. The number of results given back. Please use get_content_all if the limit exceeds 1000 rows.
offset	integer. Offset for the list of matches.
sort	character. Sort search result by various fields. For example: sort=release_date asc, uuid desc.
begin_date	character. Begin date - Restricts responses to results with publication dates of the date specified or later. In the form YYYYMMDD.
end_date	character. End date - Restricts responses to results with publication dates of the date specified or earlier. In the form YYYYMMDD.
api_key	character. The personal api code. To request an API key see: http://developer.zeit.de/quickstart/ This parameter is by default set to the R Environment.

Details

get_content is the function, which interacts directly with the ZEIT Online API. I only used the content endpoint for this package. There are further endpoints (e.g. /author, /product) not included into this package to further specify the search if needed. The whole list of possible endpoints can be accessed here <http://developer.zeit.de/docs/>.

Value

A list including articles and meta information about the query.

Author(s)

Jan Dix <<jan.dix@uni-konstanz.de>>

References

<http://developer.zeit.de>

See Also

[get_content_all](#)

Examples

```
## Not run:  
get_content(query = "Merkel")  
  
## End(Not run)
```

get_content_all	<i>Content endpoint (all)</i>
-----------------	-------------------------------

Description

Exposes a search in the ZEIT online archive on the content endpoint and returns results for the given query. Performs multiple queries if limit exceeds 1000 rows.

Usage

```
get_content_all(query, timeout = 2, begin_date = NULL,  
               end_date = NULL, api_key = Sys.getenv("ZEIT_ONLINE_KEY"))
```

Arguments

query	character. Search query term.
timeout	integer. Seconds to wait between queries.
begin_date	begin_date character. Begin date - Restricts responses to results with publication dates of the date specified or later. In the form YYYYMMDD.
end_date	character. End date - Restricts responses to results with publication dates of the date specified or earlier. In the form YYYYMMDD.
api_key	api_key character. The personal api code. To request an API key see: http://developer.zeit.de/quickstart/ This parameter is by default set to the R Environment.

Details

get_content is the function, which interacts directly with the ZEIT Online API. I only used the content endpoint for this package. There are further endpoints (e.g. /author, /product) not included into this package to further specify the search if needed. The whole list of possible endpoints can be accessed here <http://developer.zeit.de/docs/>.

Value

A list including articles and meta information about the query.

Author(s)

Jan Dix <<jan.dix@uni-konstanz.de>>

References

<http://developer.zeit.de>

See Also

[get_content](#)

Examples

```
## Not run:  
get_content(query = "Merkel")  
  
## End(Not run)
```

rzeit2

Client for the ZEIT ONLINE Content API

Description

Interface to gather newspaper articles from DIE ZEIT and ZEIT ONLINE, based on a multilevel query. A personal API key is required for usage.

References

<http://developer.zeit.de>

See Also

[get_content](#) to expose a search in the ZEIT online archive, [get_content_all](#) to get all results using [get_content](#), [get_client](#) to get client information

sentiment_example

Sentiment scores for 103 ZEIT ONLINE articles

Description

The dataset contains 103 articles returned by a query using the keyword "Merkel" between 01st May and 31st May 2018. The sentiment scores are calculated using the Sentiment Worschatz dictionary.

Usage

```
sentiment_example
```

Format

A data frame with 103 rows and 3 variables:

url the url of the article

date the release date of the article

score the calculated sentiment score

See Also

[senti_ws](#)

senti_ws	<i>SentimentWortschatz (SentiWS)</i>
----------	--------------------------------------

Description

SentiWS is a publicly available German-language resource for sentiment analysis, opinion mining etc. It lists positive and negative polarity bearing words weighted within the interval of [-1; 1].

Usage

```
senti_ws
```

Format

A data frame with 3468 rows and 2 variables:

word word

score score of the respective word

Source

R. Remus, U. Quasthoff & G. Heyer: SentiWS - a Publicly Available German-language Resource for Sentiment Analysis. In: Proceedings of the 7th International Language Ressources and Evaluation (LREC'10), pp. 1168-1171, 2010 <http://wortschatz.uni-leipzig.de/en/download/>

set_api_key	<i>Set api key to the .Renviron</i>
-------------	-------------------------------------

Description

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

Usage

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

Arguments

api_key character. The personal api code. To request an API key see: <http://developer.rzeit.de/quickstart/>

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

Value

None.

Author(s)

Jan Dix <<jan.dix@uni-konstanz.de>>

Examples

```
## Not run:  
# this is not an actual api key  
api_key <- "5t5yno5qqkufxis5q2vzx26vxq2hqe9"  
set_api_key(api_key, tempdir())  
  
## End(Not run)
```

Index

*Topic **datasets**

senti_ws, 9

sentiment_example, 8

get_article_comments, 2

get_article_images, 3

get_article_text, 4

get_client, 5, 8

get_content, 5, 7, 8

get_content_all, 6, 7, 8

rzeit2, 8

rzeit2-package (rzeit2), 8

senti_ws, 8, 9

sentiment_example, 8

set_api_key, 9