

Package ‘tidywikidatar’

September 14, 2021

Title Explore 'Wikidata' Through Tidy Data Frames

Version 0.4.2

Description Query 'Wikidata' API <https://www.wikidata.org/wiki/Wikidata:Main_Page> with ease, get tidy data frames in response, and cache data in a local 'SQLite' database.

License MIT + file LICENSE

Encoding UTF-8

RoxygenNote 7.1.2

Imports magrittr, dplyr, tidyr, WikidataR, stringr, glue, DBI, RSQLite, tibble, purrr, usethis, WikidataQueryServiceR, fs, rlang (>= 0.1.2), progress, jsonlite

Suggests spelling, roxygen2, testthat (>= 3.0.0), knitr, rmarkdown

Config/testthat/edition 3

Language en-US

URL <https://edjnet.github.io/tidywikidatar/>

VignetteBuilder knitr

NeedsCompilation no

Author Giorgio Comai [aut, cre, cph] (<<https://orcid.org/0000-0002-0515-9542>>), EDJNet [fnd]

Maintainer Giorgio Comai <g@giorgiocomai.eu>

Repository CRAN

Date/Publication 2021-09-14 16:20:02 UTC

R topics documented:

tw_check_cache	3
tw_check_cached_items	3
tw_check_cache_folder	4
tw_check_qid	5
tw_check_search	6
tw_connect_to_cache	7

tw_create_cache_folder	8
tw_disable_cache	8
tw_disconnect_from_cache	9
tw_enable_cache	10
tw_filter	10
tw_filter_first	11
tw_filter_people	12
tw_get	13
tw_get_cached_item	14
tw_get_cached_qualifiers	15
tw_get_cached_search	17
tw_get_cache_file	18
tw_get_cache_table_name	18
tw_get_description	19
tw_get_field	20
tw_get_id_of_wikipedia_page	21
tw_get_image	22
tw_get_image_metadata	23
tw_get_image_metadata_single	24
tw_get_image_same_length	26
tw_get_label	27
tw_get_links_from_wikipedia_page	29
tw_get_property	30
tw_get_property_description	31
tw_get_property_label	32
tw_get_property_same_length	33
tw_get_property_with_details	35
tw_get_property_with_details_single	36
tw_get_qualifiers	37
tw_get_qualifiers_single	38
tw_get_single	39
tw_get_wikipedia	40
tw_get_wikipedia_base_api_url	41
tw_label	42
tw_query	43
tw_search	44
tw_search_item	45
tw_search_property	47
tw_search_single	48
tw_set_cache_folder	49
tw_set_language	50
tw_write_item_to_cache	50
tw_write_qualifiers_to_cache	52
tw_write_search_to_cache	53

tw_check_cache	<i>Check caching status in the current session, and override it upon request</i>
----------------	--

Description

Mostly used internally in functions, exported for reference.

Usage

```
tw_check_cache(cache = NULL)
```

Arguments

cache	Defaults to NULL. If NULL, checks current cache settings. If given, returns given value, ignoring cache.
-------	--

Value

Either TRUE or FALSE, depending on current cache settings.

Examples

```
if (interactive()) {  
  tw_check_cache()  
}
```

tw_check_cached_items	<i>Check if given items are present in cache</i>
-----------------------	--

Description

Check if given items are present in cache

Usage

```
tw_check_cached_items(  
  id,  
  language = tidywikidatar::tw_get_language(),  
  cache_connection = NULL,  
  disconnect_db = TRUE  
)
```

Arguments

id	A character vector. Each element must start with Q, and correspond to a Wiki-data identifier.
language	Defaults to language set with <code>tw_set_language()</code> ; if not set, "en". Use "all_available" to keep all languages. For available language values, see https://www.wikidata.org/wiki/Help:Wikimedia_
cache_connection	Defaults to NULL. If NULL, and caching is enabled, <code>tidywikidatar</code> will use a local sqlite database. A custom connection to other databases can be given (see vignette caching for details).
disconnect_db	Defaults to TRUE. If FALSE, leaves the connection to cache open.

Value

A character vector with IDs of items present in cache. If no item found in cache, returns NULL.

Examples

```
tw_set_cache_folder(path = tempdir())
tw_enable_cache()
tw_create_cache_folder(ask = FALSE)

# add three items to local cache
invisible(tw_get(id = "Q180099", language = "en"))
invisible(tw_get(id = "Q228822", language = "en"))
invisible(tw_get(id = "Q184992", language = "en"))

# check if these other items are in cache
items_in_cache <- tw_check_cached_items(
  id = c(
    "Q180099",
    "Q228822",
    "Q76857"
  ),
  language = "en"
)
# it should return only the two items from the current list of id
# but not other item already in cache
items_in_cache
```

`tw_check_cache_folder` Checks if cache folder exists, if not returns an informative message

Description

Checks if cache folder exists, if not returns an informative message

Usage

```
tw_check_cache_folder()
```

Value

If the cache folder exists, returns TRUE. Otherwise throws an error.

Examples

```
# If cache folder does not exist, it throws an error
tryCatch(tw_check_cache_folder(),
  error = function(e) {
    return(e)
  }
)

# Create cache folder
tw_set_cache_folder(path = fs::path(
  tempdir(),
  "tw_cache_folder"
))
tw_create_cache_folder(ask = FALSE)

tw_check_cache_folder()
```

tw_check_qid

Ensures that input appears to be a valid Wikidata id

Description

Mostly used internally by other functions.

Usage

```
tw_check_qid(id)
```

Arguments

id A character vector of one or more Wikidata id.

Value

A character vector with only strings appearing to be Wikidata identifiers; possibly shorter than input

Examples

```
tw_check_qid(id = c("Q180099", "q228822", "Not an id", "00180099", NA))
```

tw_check_search	<i>Checks if an input is a search; if not, it tries to return a search</i>
-----------------	--

Description

Mostly used as a convenience function inside other functions to have consistent inputs.

Usage

```
tw_check_search(
  search,
  type = "item",
  language = tidywikidatar::tw_get_language(),
  limit = 10,
  include_search = FALSE,
  wait = 0,
  cache = NULL,
  overwrite_cache = FALSE,
  cache_connection = NULL,
  disconnect_db = TRUE
)
```

Arguments

search	A string to be searched in Wikidata
type	Defaults to "item". Either "item" or "property".
language	Language to be used for the search. Can be set once per session with <code>tw_set_language()</code> . If not set, defaults to "en". For a full list, see https://www.wikidata.org/wiki/Help:Wikimedia_language_co
limit	Maximum numbers of responses to be given.
include_search	Logical, defaults to FALSE. If TRUE, the search is returned as an additional column.
wait	In seconds, defaults to 0. Time to wait between queries to Wikidata. If data are cached locally, wait time is not applied. If you are running many queries systematically you may want to add some waiting time between queries.
cache	Defaults to NULL. If given, it should be given either TRUE or FALSE. Typically set with <code>tw_enable_cache()</code> or <code>tw_disable_cache()</code> .
overwrite_cache	Defaults to FALSE. If TRUE, overwrites cache.
cache_connection	Defaults to NULL. If NULL, and caching is enabled, <code>tidywikidatar</code> will use a local sqlite database. A custom connection to other databases can be given (see vignette caching for details).
disconnect_db	Defaults to TRUE. If FALSE, leaves the connection to cache open.

Value

A data frame with three columns, id, label, and description, filtered by the above criteria.

Examples

```
# The following two lines should give the same result.

tw_check_search("Sylvia Pankhurst")
tw_check_search(tw_search("Sylvia Pankhurst"))
```

```
tw_connect_to_cache    Return a connection to be used for caching
```

Description

Return a connection to be used for caching

Usage

```
tw_connect_to_cache(connection = NULL, language = NULL)
```

Arguments

connection	Defaults to NULL. If NULL, uses local SQLite database. If given, must be a connection object (see example)
language	Defaults to language set with <code>tw_set_language()</code> ; if not set, "en". Use "all_available" to keep all languages. For available language values, see https://www.wikidata.org/wiki/Help:Wikimedia

Value

A connection object.

Examples

```
if (interactive()) {
  cache_connection <- DBI::dbConnect(
    RSQLite::SQLite(), # or e.g. odbc::odbc(),
    Driver = ":memory:", # or e.g. "MariaDB",
    Host = "localhost",
    database = "example_db",
    UID = "example_user",
    PWD = "example_pwd"
  )
  tw_connect_to_cache(cache_connection)
}
```

`tw_create_cache_folder`*Creates the base cache folder where tidywikidatar caches data.*

Description

Creates the base cache folder where tidywikidatar caches data.

Usage

```
tw_create_cache_folder(ask = TRUE)
```

Arguments

ask	Logical, defaults to TRUE. If FALSE, and cache folder does not exist, it just creates it without asking (useful for interactive sessions).
-----	--

Value

Nothing, used for its side effects.

Examples

```
if (interactive()) {  
  tw_create_cache_folder()  
}
```

`tw_disable_cache`*Disable caching for the current session*

Description

Disable caching for the current session

Usage

```
tw_disable_cache()
```

Value

Nothing, used for its side effects.

Examples

```
if (interactive()) {
  tw_disable_cache()
}
```

tw_disconnect_from_cache

Ensure that connection to cache is disconnected consistently

Description

Ensure that connection to cache is disconnected consistently

Usage

```
tw_disconnect_from_cache(
  cache = NULL,
  cache_connection = NULL,
  disconnect_db = TRUE,
  language = tidywikidatar::tw_get_language()
)
```

Arguments

cache Defaults to NULL. If given, it should be given either TRUE or FALSE. Typically set with `tw_enable_cache()` or `tw_disable_cache()`.

cache_connection Defaults to NULL. If NULL, and caching is enabled, `tidywikidatar` will use a local sqlite database. A custom connection to other databases can be given (see vignette caching for details).

disconnect_db Defaults to TRUE. If FALSE, leaves the connection to cache open.

language Defaults to language set with `tw_set_language()`; if not set, "en". Use "all_available" to keep all languages. For available language values, see <https://www.wikidata.org/wiki/Help:Wikimedia>

Value

Nothing, used for its side effects.

Examples

```
tw_get(
  id = c("Q180099"),
  language = "en"
)
tw_disconnect_from_cache()
```

tw_enable_cache	<i>Enable caching for the current session</i>
-----------------	---

Description

Enable caching for the current session

Usage

```
tw_enable_cache()
```

Value

Nothing, used for its side effects.

Examples

```
if (interactive()) {  
    tw_enable_cache()  
}
```

tw_filter	<i>Filter search result and keep only items with matching property and Q id</i>
-----------	---

Description

Filter search result and keep only items with matching property and Q id

Usage

```
tw_filter(  
    search,  
    p,  
    q,  
    language = tidywikidatar::tw_get_language(),  
    limit = 10,  
    include_search = FALSE,  
    wait = 0,  
    cache = NULL,  
    overwrite_cache = FALSE,  
    cache_connection = NULL,  
    disconnect_db = TRUE  
)
```

Arguments

search	A data frame generated by <code>tw_search()</code> , or a search query. If a data frame is given, language and limits are ignore.
p	A character vector of length 1, a property. Must always start with the capital letter "P", e.g. "P31" for "instance of".
q	A character vector of length 1, a wikidata id. Must always start with the capital letter "Q", e.g. "Q5" for "human being".
language	Language to be used for the search. Can be set once per session with <code>tw_set_language()</code> . If not set, defaults to "en". For a full list, see https://www.wikidata.org/wiki/Help:Wikimedia_language_codes .
limit	Maximum numbers of responses to be given.
include_search	Logical, defaults to FALSE. If TRUE, the search is returned as an additional column.
wait	In seconds, defaults to 0. Time to wait between queries to Wikidata. If data are cached locally, wait time is not applied. If you are running many queries systematically you may want to add some waiting time between queries.
cache	Defaults to NULL. If given, it should be given either TRUE or FALSE. Typically set with <code>tw_enable_cache()</code> or <code>tw_disable_cache()</code> .
overwrite_cache	Defaults to FALSE. If TRUE, overwrites cache.
cache_connection	Defaults to NULL. If NULL, and caching is enabled, <code>tidywikidatar</code> will use a local sqlite database. A custom connection to other databases can be given (see vignette caching for details).
disconnect_db	Defaults to TRUE. If FALSE, leaves the connection to cache open.

Value

A data frame with three columns, `id`, `label`, and `description`, filtered by the above criteria.

Examples

```
tw_search(search = "Margaret Mead", limit = 3) %>%
  tw_filter(p = "P31", q = "Q5")
```

<code>tw_filter_first</code>	<i>Filter search result and keep only and keep only the first match</i>
------------------------------	---

Description

Same as `tw_filter()`, but consistently returns data frames with a single row.

Usage

```
tw_filter_first(
  search,
  p,
  q,
  language = "en",
  limit = 10,
  wait = 0,
  cache = NULL
)
```

Arguments

search	A data frame generated by tw_search(), or a search query. If a data frame is given, language and limits are ignore.
p	A character vector of length 1, a property. Must always start with the capital letter "P", e.g. "P31" for "instance of".
q	A character vector of length 1, a wikidata id. Must always start with the capital letter "Q", e.g. "Q5" for "human being".
language	Language to be used for the search.
limit	Maximum numbers of responses to be given.
wait	In seconds, defaults to 0. Time to wait between queries to Wikidata. If data are cached locally, wait time is not applied. If you are running many queries systematically you may want to add some waiting time between queries.
cache	Defaults to NULL. If given, it should be given either TRUE or FALSE. Typically set with tw_enable_cache() or tw_disable_cache()

Value

A data frame with one row and three columns, id, label, and description, filtered by the above criteria.

Examples

```
tw_search("Margaret Mead") %>%
  tw_filter_first(p = "P31", q = "Q5")
```

tw_filter_people	<i>Filter search result and keep only people</i>
------------------	--

Description

A wrapper of tw_filter() that defaults to keep only "instance of" (P31) "human being" (Q5).

Usage

```
tw_filter_people(
  search,
  language = "en",
  limit = 10,
  stop_at_first = TRUE,
  wait = 0
)
```

Arguments

search	A data frame generated by <code>tw_search()</code> , or a search query. If a data frame is given, language and limits are ignore.
language	Language to be used for the search.
limit	Maximum numbers of responses to be given.
stop_at_first	Logical, defaults to TRUE. If TRUE, returns only the first match from the search that satisfies the criteria.
wait	In seconds, defaults to 0. Time to wait between queries to Wikidata. If data are cached locally, wait time is not applied. If you are running many queries systematically you may want to add some waiting time between queries.

Value

A data frame with three columns, `id`, `label`, and `description`; all rows refer to a human being.

Examples

```
tw_search("Ruth Benedict")

tw_search("Ruth Benedict") %>%
  tw_filter_people()
```

tw_get

Return (most) information from a Wikidata item in a tidy format

Description

Return (most) information from a Wikidata item in a tidy format

Usage

```
tw_get(
  id,
  language = tidywikidatar::tw_get_language(),
  cache = NULL,
  overwrite_cache = FALSE,
```

```

    cache_connection = NULL,
    disconnect_db = TRUE,
    wait = 0
  )

```

Arguments

id	A character vector, must start with Q, e.g. "Q180099" for the anthropologist Margaret Mead. Can also be a data frame of one row, typically generated with <code>tw_search()</code> or a combination of <code>tw_search()</code> and <code>tw_filter_first()</code> .
language	Defaults to language set with <code>tw_set_language()</code> ; if not set, "en". Use "all_available" to keep all languages. For available language values, see https://www.wikidata.org/wiki/Help:Wikimedia .
cache	Defaults to NULL. If given, it should be given either TRUE or FALSE. Typically set with <code>tw_enable_cache()</code> or <code>tw_disable_cache()</code> .
overwrite_cache	Logical, defaults to FALSE. If TRUE, it overwrites the table in the local sqlite database. Useful if the original Wikidata object has been updated.
cache_connection	Defaults to NULL. If NULL, and caching is enabled, <code>tidywikidatar</code> will use a local sqlite database. A custom connection to other databases can be given (see vignette caching for details).
disconnect_db	Defaults to TRUE. If FALSE, leaves the connection to cache open.
wait	In seconds, defaults to 0. Time to wait between queries to Wikidata. If data are cached locally, wait time is not applied. If you are running many queries systematically you may want to add some waiting time between queries.

Value

A data.frame (a tibble) with three columns (id, property, and value).

Examples

```

tw_get(
  id = c("Q180099", "Q228822"),
  language = "en"
)

```

tw_get_cached_item	<i>Retrieve cached item</i>
--------------------	-----------------------------

Description

Retrieve cached item

Usage

```
tw_get_cached_item(
  id,
  language = tidywikidatar::tw_get_language(),
  cache_connection = NULL,
  disconnect_db = TRUE
)
```

Arguments

id A character vector, must start with Q, e.g. "Q180099" for the anthropologist Margaret Mead. Can also be a data frame of one row, typically generated with `tw_search()` or a combination of `tw_search()` and `tw_filter_first()`.

language Defaults to language set with `tw_set_language()`; if not set, "en". Use "all_available" to keep all languages. For available language values, see <https://www.wikidata.org/wiki/Help:Wikimedia>

cache_connection Defaults to NULL. If NULL, and caching is enabled, `tidywikidatar` will use a local sqlite database. A custom connection to other databases can be given (see vignette caching for details).

disconnect_db Defaults to TRUE. If FALSE, leaves the connection open.

Value

If data present in cache, returns a data frame with cached data.

Examples

```
tw_set_cache_folder(path = tempdir())
tw_enable_cache()
tw_create_cache_folder(ask = FALSE)

df_from_api <- tw_get(id = "Q180099", language = "en")

df_from_cache <- tw_get_cached_item(
  id = "Q180099",
  language = "en"
)
```

tw_get_cached_qualifiers

Retrieve cached qualifier

Description

Retrieve cached qualifier

Usage

```
tw_get_cached_qualifiers(
  id,
  p,
  language = tidywikidatar::tw_get_language(),
  cache_connection = NULL,
  disconnect_db = TRUE
)
```

Arguments

id	A character vector, must start with Q, e.g. "Q180099" for the anthropologist Margaret Mead. Can also be a data frame of one row, typically generated with <code>tw_search()</code> or a combination of <code>tw_search()</code> and <code>tw_filter_first()</code> .
p	A character vector of length 1, a property. Must always start with the capital letter "P", e.g. "P31" for "instance of".
language	Defaults to language set with <code>tw_set_language()</code> ; if not set, "en". Use "all_available" to keep all languages. For available language values, see https://www.wikidata.org/wiki/Help:Wikimedia
cache_connection	Defaults to NULL. If NULL, and caching is enabled, <code>tidywikidatar</code> will use a local sqlite database. A custom connection to other databases can be given (see vignette caching for details).
disconnect_db	Defaults to TRUE. If FALSE, leaves the connection open.

Value

If data present in cache, returns a data frame with cached data.

Examples

```
tw_set_cache_folder(path = tempdir())
tw_enable_cache()
tw_create_cache_folder(ask = FALSE)

df_from_api <- tw_get_qualifiers(id = "Q180099", p = "P26", language = "en")

df_from_cache <- tw_get_cached_qualifiers(
  id = "Q180099",
  p = "P26",
  language = "en"
)

df_from_cache
```

tw_get_cached_search *Retrieve cached search*

Description

Retrieve cached search

Usage

```
tw_get_cached_search(
  search,
  type = "item",
  language = tidywikidatar::tw_get_language(),
  include_search = FALSE,
  cache_connection = NULL,
  disconnect_db = TRUE
)
```

Arguments

search	A string to be searched in Wikidata
type	Defaults to "item". Either "item" or "property".
language	Defaults to language set with <code>tw_set_language()</code> ; if not set, "en". Use "all_available" to keep all languages. For available language values, see https://www.wikidata.org/wiki/Help:Wikimedia_
include_search	Logical, defaults to FALSE. If TRUE, the search is returned as an additional column.
cache_connection	Defaults to NULL. If NULL, and caching is enabled, tidywikidatar will use a local sqlite database. A custom connection to other databases can be given (see vignette caching for details).
disconnect_db	Defaults to TRUE. If FALSE, leaves the connection to cache open.

Value

If data present in cache, returns a data frame with cached data.

Examples

```
tw_set_cache_folder(path = tempdir())
tw_enable_cache()
tw_create_cache_folder(ask = FALSE)

search_from_api <- tw_search("Sylvia Pankhurst")
search_from_api
```

```
df_from_cache <- tw_get_cached_search("Sylvia Pankhurst")
df_from_cache
```

```
tw_get_cache_file      Gets location of cache file
```

Description

Gets location of cache file

Usage

```
tw_get_cache_file(type = "item", language = tidywikidatar::tw_get_language())
```

Arguments

type	Defaults to "item". Type of cache file to output. Values typically used by tidywikidatar include "item", "search", and "qualifier".
language	Defaults to language set with tw_set_language(); if not set, "en". Use "all_available" to keep all languages. For available language values, see https://www.wikidata.org/wiki/Help:Wikimedia_

Value

A character vector of length one with location of item cache file.

Examples

```
tw_set_cache_folder(path = tempdir())
sqlite_cache_file_location <- tw_get_cache_file() # outputs location of cache file
```

```
tw_get_cache_table_name
      Gets name of table inside the database
```

Description

Gets name of table inside the database

Usage

```
tw_get_cache_table_name(
  type = "item",
  language = tidywikidatar::tw_get_language()
)
```

Arguments

type	Defaults to "item". Type of cache file to output. Values typically used by tidywikidatar include "item", "search", and "qualifier".
language	Defaults to language set with tw_set_language(); "en" if not set. Used to limit the data to be cached. Use "all_available" to keep all data. For available values, see https://www.wikidata.org/wiki/Help:Wikimedia_language_codes/lists/all

Value

A character vector of length one with the name of the relevant table in the cache file.

Examples

```
# outputs name of table used in of cache file
tw_get_cache_table_name(type = "item", language = "en")
```

tw_get_description	<i>Get Wikidata description in given language</i>
--------------------	---

Description

Get Wikidata description in given language

Usage

```
tw_get_description(
  id,
  language = tidywikidatar::tw_get_language(),
  id_df = NULL,
  cache = NULL,
  overwrite_cache = FALSE,
  cache_connection = NULL,
  disconnect_db = TRUE,
  wait = 0
)
```

Arguments

id	A character vector, must start with Q, e.g. "Q254" for Wolfgang Amadeus Mozart
language	Defaults to language set with tw_set_language(); if not set, "en". Use "all_available" to keep all languages. For available language values, see https://www.wikidata.org/wiki/Help:Wikimedia_
id_df	Default to NULL. If given, it should be a dataframe typically generated with tw_get_(), and is used instead of calling Wikidata or using SQLite cache. Ignored when id is of length more than one.

cache	Defaults to NULL. If given, it should be given either TRUE or FALSE. Typically set with <code>tw_enable_cache()</code> or <code>tw_disable_cache()</code> .
overwrite_cache	Logical, defaults to FALSE. If TRUE, it overwrites the table in the local sqlite database. Useful if the original Wikidata object has been updated.
cache_connection	Defaults to NULL. If NULL, and caching is enabled, <code>tidywikidatar</code> will use a local sqlite database. A custom connection to other databases can be given (see vignette caching for details).
disconnect_db	Defaults to TRUE. If FALSE, leaves the connection to cache open.
wait	In seconds, defaults to 0. Time to wait between queries to Wikidata. If data are cached locally, wait time is not applied. If you are running many queries systematically you may want to add some waiting time between queries.

Value

A character vector of length 1, with the Wikidata description in the requested language.

Examples

```
tw_get_description(
  id = c(
    "Q180099",
    "Q228822"
  ),
  language = "en"
)
```

tw_get_field	<i>Gets a field such a label or description from a dataframe typically generated with tw_get()</i>
--------------	--

Description

Gets a field such a label or description from a dataframe typically generated with `tw_get()`

Usage

```
tw_get_field(df, field, id, language = tidywikidatar::tw_get_language())
```

Arguments

df	A data frame typically generated with <code>tw_get()</code> . It should include data for the id included in the dedicated parameter.
field	A character vector of length one. Typically, either "label" or "description".
id	A character vector, typically of Wikidata identifiers. The output will be of the same length and in the same order as the identifiers provided with this parameter.
language	Defaults to language set with <code>tw_set_language()</code> ; if not set, "en". Use "all_available" to keep all languages. For available language values, see https://www.wikidata.org/wiki/Help:Wikimedia_

Value

A character vector of the same length, with data in the same order, as id.

Examples

```
tw_get("Q180099") %>%  
  tw_get_field(field = "label", id = "Q180099")
```

tw_get_id_of_wikipedia_page

Gets the Wikidata id of a Wikipedia page

Description

Gets the Wikidata id of a Wikipedia page

Usage

```
tw_get_id_of_wikipedia_page(url = NULL, title = NULL, language = NULL)
```

Arguments

url	Full URL to a Wikipedia page. If given, title and language can be left empty.
title	Title of a Wikipedia page or final parts of its url. If given, url can be left empty, but language must be provided.
language	Two-letter language code used to define the Wikipedia version to use. If url given, this can be left empty.

Value

A character vector of Wikidata identifiers.

Examples

```
if (interactive()) {  
  tw_get_id_of_wikipedia_page(title = "Margaret Mead", language = "en")  
}
```

tw_get_image	<i>Get image from Wikimedia Commons</i>
--------------	---

Description

Please consult the relevant documentation for reusing content outside Wikimedia: <https://commons.wikimedia.org/wiki/Com>

Usage

```
tw_get_image(
  id,
  format = "filename",
  width = NULL,
  language = tidywikidatar::tw_get_language(),
  id_df = NULL,
  cache = NULL,
  overwrite_cache = FALSE,
  cache_connection = NULL,
  disconnect_db = TRUE,
  wait = 0
)
```

Arguments

id	A character vector of length 1, must start with Q, e.g. "Q254" for Wolfgang Amadeus Mozart.
format	A character vector, defaults to 'filename'. If set to 'commons', outputs the link to the Wikimedia Commons page. If set to "embed", outputs a link that can be used to embed.
width	A numeric value, defaults to NULL, relevant only if format is set to 'embed'. If not given, defaults to full resolution image.
language	Needed for caching, defaults to language set with tw_set_language(); if not set, "en". Use "all_available" to keep all languages. For available language values, see https://www.wikidata.org/wiki/Help:Wikimedia_language_codes/lists/all
id_df	Default to NULL. If given, it should be a dataframe typically generated with tw_get_(), and is used instead of calling Wikidata or using SQLite cache. Ignored when id is of length more than one.
cache	Defaults to NULL. If given, it should be given either TRUE or FALSE. Typically set with tw_enable_cache() or tw_disable_cache().
overwrite_cache	Logical, defaults to FALSE. If TRUE, it overwrites the table in the local sqlite database. Useful if the original Wikidata object has been updated.
cache_connection	Defaults to NULL. If NULL, and caching is enabled, tidywikidatar will use a local sqlite database. A custom connection to other databases can be given (see vignette caching for details).

`disconnect_db` Defaults to TRUE. If FALSE, leaves the connection to cache open.

`wait` In seconds, defaults to 0. Time to wait between queries to Wikidata. If data are cached locally, wait time is not applied. If you are running many queries systematically you may want to add some waiting time between queries.

Value

A data frame of two columns, `id` and `image`, corresponding to reference to the image in the requested format.

Examples

```
tw_get_image("Q180099",
  format = "filename"
)

tw_get_image("Q180099",
  format = "commons"
)

tw_get_image("Q180099",
  format = "embed",
  width = 300
)
```

`tw_get_image_metadata` *Get metadata for images from Wikimedia Commons*

Description

Please consult the relevant documentation for reusing content outside Wikimedia: https://commons.wikimedia.org/wiki/Commons:Reusing_content_outside_Wikimedia

Usage

```
tw_get_image_metadata(
  id,
  image_filename = NULL,
  only_first = TRUE,
  language = tidywikidatar::tw_get_language(),
  id_df = NULL,
  cache = NULL,
  overwrite_cache = FALSE,
  cache_connection = NULL,
  disconnect_db = TRUE,
  wait = 0
)
```

Arguments

id	A character vector of length 1, must start with Q, e.g. "Q254" for Wolfgang Amadeus Mozart.
image_filename	Defaults to NULL. If NULL, image_filename is obtained from the Wikidata id. If given, must be of the same length as id.
only_first	Defaults to TRUE. If TRUE, returns metadata only for the first image associated with a given Wikidata id. If FALSE, returns all images available.
language	Needed for caching, defaults to language set with tw_set_language(); if not set, "en". Use "all_available" to keep all languages. For available language values, see https://www.wikidata.org/wiki/Help:Wikimedia_language_codes/lists/all
id_df	Default to NULL. If given, it should be a dataframe typically generated with tw_get_(), and is used instead of calling Wikidata or using SQLite cache. Ignored when id is of length more than one.
cache	Defaults to NULL. If given, it should be given either TRUE or FALSE. Typically set with tw_enable_cache() or tw_disable_cache().
overwrite_cache	Logical, defaults to FALSE. If TRUE, it overwrites the table in the local sqlite database. Useful if the original Wikidata object has been updated.
cache_connection	Defaults to NULL. If NULL, and caching is enabled, tidywikidatar will use a local sqlite database. A custom connection to other databases can be given (see vignette caching for details).
disconnect_db	Defaults to TRUE. If FALSE, leaves the connection to cache open.
wait	In seconds, defaults to 0. Time to wait between queries to Wikidata. If data are cached locally, wait time is not applied. If you are running many queries systematically you may want to add some waiting time between queries.

Value

A character vector, corresponding to reference to the image in the requested format.

Examples

```
if (interactive()) {
  tw_get_image_metadata("Q180099")
}
```

tw_get_image_metadata_single

Get metadata for images from Wikimedia Commons

Description

Please consult the relevant documentation for reusing content outside Wikimedia: https://commons.wikimedia.org/wiki/Commons:Reusing_content_outside_Wikimedia

Usage

```
tw_get_image_metadata_single(
    id,
    image_filename = NULL,
    only_first = TRUE,
    language = tidywikidatar::tw_get_language(),
    id_df = NULL,
    cache = NULL,
    overwrite_cache = FALSE,
    read_cache = TRUE,
    cache_connection = NULL,
    disconnect_db = TRUE,
    wait = 0
)
```

Arguments

<code>id</code>	A character vector of length 1, must start with Q, e.g. "Q254" for Wolfgang Amadeus Mozart.
<code>image_filename</code>	Defaults to NULL. If NULL, <code>image_filename</code> is obtained from the Wikidata id. If given, must be of the same length as <code>id</code> .
<code>only_first</code>	Defaults to TRUE. If TRUE, returns metadata only for the first image associated with a given Wikidata id. If FALSE, returns all images available.
<code>language</code>	Needed for caching, defaults to language set with <code>tw_set_language()</code> ; if not set, "en". Use "all_available" to keep all languages. For available language values, see https://www.wikidata.org/wiki/Help:Wikimedia_language_codes/lists/all
<code>id_df</code>	Default to NULL. If given, it should be a dataframe typically generated with <code>tw_get_()</code> , and is used instead of calling Wikidata or using SQLite cache. Ignored when <code>id</code> is of length more than one.
<code>cache</code>	Defaults to NULL. If given, it should be given either TRUE or FALSE. Typically set with <code>tw_enable_cache()</code> or <code>tw_disable_cache()</code> .
<code>overwrite_cache</code>	Logical, defaults to FALSE. If TRUE, it overwrites the table in the local sqlite database. Useful if the original Wikidata object has been updated.
<code>read_cache</code>	Logical, defaults to TRUE. Mostly used internally to prevent checking if an item is in cache if it is already known that it is not in cache.
<code>cache_connection</code>	Defaults to NULL. If NULL, and caching is enabled, <code>tidywikidatar</code> will use a local sqlite database. A custom connection to other databases can be given (see vignette caching for details).
<code>disconnect_db</code>	Defaults to TRUE. If FALSE, leaves the connection to cache open.
<code>wait</code>	In seconds, defaults to 0. Time to wait between queries to Wikidata. If data are cached locally, wait time is not applied. If you are running many queries systematically you may want to add some waiting time between queries.

Value

A character vector, corresponding to reference to the image in the requested format.

Examples

```
if (interactive()) {
  tw_get_image_metadata_single("Q180099")
}
```

```
tw_get_image_same_length
```

Get image from Wikimedia Commons

Description

Please consult the relevant documentation for reusing content outside Wikimedia: https://commons.wikimedia.org/wiki/Commons:Reusing_content_outside_Wikimedia

Usage

```
tw_get_image_same_length(
  id,
  format = "filename",
  as_tibble = FALSE,
  only_first = TRUE,
  width = NULL,
  language = tidywikidatar::tw_get_language(),
  id_df = NULL,
  cache = NULL,
  overwrite_cache = FALSE,
  cache_connection = NULL,
  disconnect_db = TRUE,
  wait = 0
)
```

Arguments

id	A character vector of length 1, must start with Q, e.g. "Q254" for Wolfgang Amadeus Mozart.
format	A character vector, defaults to 'filename'. If set to 'commons', outputs the link to the Wikimedia Commons page. If set to "embed", outputs a link that can be used to embed.
as_tibble	Defaults to FALSE. If TRUE, returns a data frame instead of a character vector.
only_first	Defaults to TRUE. If TRUE, returns only the first image associated with a given Wikidata id. If FALSE, returns all images available.
width	A numeric value, defaults to NULL, relevant only if format is set to 'embed'. If not given, defaults to full resolution image.

language	Needed for caching, defaults to language set with <code>tw_set_language()</code> ; if not set, "en". Use "all_available" to keep all languages. For available language values, see https://www.wikidata.org/wiki/Help:Wikimedia_language_codes/lists/all
id_df	Default to NULL. If given, it should be a dataframe typically generated with <code>tw_get_()</code> , and is used instead of calling Wikidata or using SQLite cache. Ignored when id is of length more than one.
cache	Defaults to NULL. If given, it should be given either TRUE or FALSE. Typically set with <code>tw_enable_cache()</code> or <code>tw_disable_cache()</code> .
overwrite_cache	Logical, defaults to FALSE. If TRUE, it overwrites the table in the local sqlite database. Useful if the original Wikidata object has been updated.
cache_connection	Defaults to NULL. If NULL, and caching is enabled, <code>tidywikidatar</code> will use a local sqlite database. A custom connection to other databases can be given (see vignette caching for details).
disconnect_db	Defaults to TRUE. If FALSE, leaves the connection to cache open.
wait	In seconds, defaults to 0. Time to wait between queries to Wikidata. If data are cached locally, wait time is not applied. If you are running many queries systematically you may want to add some waiting time between queries.

Value

A character vector, corresponding to reference to the image in the requested format.

Examples

```
tw_get_image_same_length("Q180099",
  format = "filename"
)

tw_get_image_same_length("Q180099",
  format = "commons"
)

tw_get_image_same_length("Q180099",
  format = "embed",
  width = 300
)
```

tw_get_label

Get Wikidata label in given language

Description

Get Wikidata label in given language

Usage

```
tw_get_label(
  id,
  language = tidywikidatar::tw_get_language(),
  id_df = NULL,
  cache = NULL,
  overwrite_cache = FALSE,
  cache_connection = NULL,
  disconnect_db = TRUE,
  wait = 0
)
```

Arguments

id	A character vector, must start with Q, e.g. "Q254" for Wolfgang Amadeus Mozart
language	Defaults to language set with tw_set_language(); if not set, "en". Use "all_available" to keep all languages. For available language values, see https://www.wikidata.org/wiki/Help:Wikimedia_
id_df	Default to NULL. If given, it should be a dataframe typically generated with tw_get_(), and is used instead of calling Wikidata or using SQLite cache. Ignored when id is of length more than one.
cache	Defaults to NULL. If given, it should be given either TRUE or FALSE. Typically set with tw_enable_cache() or tw_disable_cache().
overwrite_cache	Logical, defaults to FALSE. If TRUE, it overwrites the table in the local sqlite database. Useful if the original Wikidata object has been updated.
cache_connection	Defaults to NULL. If NULL, and caching is enabled, tidywikidatar will use a local sqlite database. A custom connection to other databases can be given (see vignette caching for details).
disconnect_db	Defaults to TRUE. If FALSE, leaves the connection to cache open.
wait	In seconds, defaults to 0. Time to wait between queries to Wikidata. If data are cached locally, wait time is not applied. If you are running many queries systematically you may want to add some waiting time between queries.

Value

A character vector of the same length as the vector of id given, with the Wikidata label in the requested language.

Examples

```
tw_get_label(
  id = c(
    "Q180099",
    "Q228822"
  )
)
```

```
),
  language = "en"
)

# If a label is not available, a NA value is returned
tw_get_label(
  id = c(
    "Q64733534",
    "Q4773904",
    "Q220480"
  ),
  language = "sc"
)
```

tw_get_links_from_wikipedia_page

Get all Wikidata id of all Wikipedia pages included in a given page

Description

Get all Wikidata id of all Wikipedia pages included in a given page

Usage

```
tw_get_links_from_wikipedia_page(url = NULL, title = NULL, language = NULL)
```

Arguments

url	Full URL to a Wikipedia page. If given, title and language can be left empty.
title	Title of a Wikipedia page or final parts of its url. If given, url can be left empty, but language must be provided.
language	Two-letter language code used to define the Wikipedia version to use. If url given, this can be left empty.

Value

A data frame (a tibble) with four columns: wikipedia_title, wikipedia_id, wikidata_id, wikidata_description.

Examples

```
if (interactive()) {
  tw_get_links_from_wikipedia_page(title = "Margaret Mead", language = "en")
}
```

tw_get_property	<i>Get Wikidata property of one or more items as a tidy data frame</i>
-----------------	--

Description

Get Wikidata property of one or more items as a tidy data frame

Usage

```
tw_get_property(
  id,
  p,
  language = tidywikidatar::tw_get_language(),
  id_df = NULL,
  cache = NULL,
  overwrite_cache = FALSE,
  cache_connection = NULL,
  disconnect_db = TRUE,
  wait = 0
)
```

Arguments

id	A character vector, must start with Q, e.g. "Q254" for Wolfgang Amadeus Mozart.
p	A character vector, a property. Must always start with the capital letter "P", e.g. "P31" for "instance of".
language	Defaults to language set with tw_set_language(); if not set, "en". Use "all_available" to keep all languages. For available language values, see https://www.wikidata.org/wiki/Help:Wikimedia
id_df	Default to NULL. If given, it should be a dataframe typically generated with tw_get_(), and is used instead of calling Wikidata or using SQLite cache. Ignored when id is of length more than one.
cache	Defaults to NULL. If given, it should be given either TRUE or FALSE. Typically set with tw_enable_cache() or tw_disable_cache().
overwrite_cache	Logical, defaults to FALSE. If TRUE, it overwrites the table in the local sqlite database. Useful if the original Wikidata object has been updated.
cache_connection	Defaults to NULL. If NULL, and caching is enabled, tidywikidatar will use a local sqlite database. A custom connection to other databases can be given (see vignette caching for details).
disconnect_db	Defaults to TRUE. If FALSE, leaves the connection to cache open.
wait	In seconds, defaults to 0. Time to wait between queries to Wikidata. If data are cached locally, wait time is not applied. If you are running many queries systematically you may want to add some waiting time between queries.

Value

A tibble, corresponding to the value for the given property. A tibble of zero rows if no relevant property found.

Examples

```
# Who were the doctoral advisors - P184 - of Margaret Mead - Q180099?
advisors <- tw_get_property(id = "Q180099", p = "P184")
advisors

# tw_get_label(advisors)

# It is also possible to get one property for many id

tw_get_property(
  id = c(
    "Q180099",
    "Q228822"
  ),
  p = "P31"
)

# Or many properties for a single id

if (interactive()) {
  tw_get_property(
    id = "Q180099",
    p = c("P21", "P31")
  )
}
```

tw_get_property_description

Get description of a Wikidata property in a given language

Description

Get description of a Wikidata property in a given language

Usage

```
tw_get_property_description(
  property,
  language = tidywikidatar::tw_get_language(),
  cache = NULL,
  overwrite_cache = FALSE,
  cache_connection = NULL,
  wait = 0
)
```

Arguments

property	A character vector of length 1, must start with P, e.g. "P31".
language	Defaults to language set with <code>tw_set_language()</code> ; if not set, "en". Use "all_available" to keep all languages. For available language values, see https://www.wikidata.org/wiki/Help:Wikimedia
cache	Defaults to NULL. If given, it should be given either TRUE or FALSE. Typically set with <code>tw_enable_cache()</code> or <code>tw_disable_cache()</code> .
overwrite_cache	Logical, defaults to FALSE. If TRUE, it overwrites the table in the local sqlite database. Useful if the original Wikidata object has been updated.
cache_connection	Defaults to NULL. If NULL, and caching is enabled, tidywikidatar will use a local sqlite database. A custom connection to other databases can be given (see vignette caching for details).
wait	In seconds, defaults to 0. Time to wait between queries to Wikidata. If data are cached locally, wait time is not applied. If you are running many queries systematically you may want to add some waiting time between queries.

Value

A character vector of length 1, with the Wikidata label in the requested language.

Examples

```
tw_get_property_description(property = "P31")
```

`tw_get_property_label` *Get label of a Wikidata property in a given language*

Description

Get label of a Wikidata property in a given language

Usage

```
tw_get_property_label(  
  property,  
  language = tidywikidatar::tw_get_language(),  
  cache = NULL,  
  overwrite_cache = FALSE,  
  cache_connection = NULL,  
  disconnect_db = TRUE,  
  wait = 0  
)
```


Arguments

property	A character vector. Each element must start with P, e.g. "P31".
language	Defaults to language set with <code>tw_set_language()</code> ; if not set, "en". Use "all_available" to keep all languages. For available language values, see https://www.wikidata.org/wiki/Help:Wikimedia_
cache	Defaults to NULL. If given, it should be given either TRUE or FALSE. Typically set with <code>tw_enable_cache()</code> or <code>tw_disable_cache()</code> .
overwrite_cache	Logical, defaults to FALSE. If TRUE, it overwrites the table in the local sqlite database. Useful if the original Wikidata object has been updated.
cache_connection	Defaults to NULL. If NULL, and caching is enabled, <code>tidywikidatar</code> will use a local sqlite database. A custom connection to other databases can be given (see vignette caching for details).
disconnect_db	Defaults to TRUE. If FALSE, leaves the connection to cache open.
wait	In seconds, defaults to 0. Time to wait between queries to Wikidata. If data are cached locally, wait time is not applied. If you are running many queries systematically you may want to add some waiting time between queries.

Value

A character vector of length 1, with the Wikidata label in the requested language.

Examples

```
tw_get_property_label(property = "P31")
```

```
tw_get_property_same_length
```

Get Wikidata property of an item as a character vector of the same length as input

Description

Get Wikidata property of an item as a character vector of the same length as input

Usage

```
tw_get_property_same_length(
  id,
  p,
  only_first = FALSE,
  language = tidywikidatar::tw_get_language(),
  id_df = NULL,
  cache = NULL,
  overwrite_cache = FALSE,
```

```

    cache_connection = NULL,
    disconnect_db = TRUE,
    wait = 0
)

```

Arguments

id	A character vector, must start with Q, e.g. "Q254" for Wolfgang Amadeus Mozart.
p	A character vector, a property. Must always start with the capital letter "P", e.g. "P31" for "instance of".
only_first	Logical, defaults to FALSE. If TRUE, it just keeps the first relevant property value for each id (or NA if none is available), and returns a character vector. Warning: this likely discards valid values, so make sure this is really what you want. If FALSE, returns a list of the same length as input, with all values for each id stored in a list if more than one is found.
language	Defaults to language set with tw_set_language(); if not set, "en". Use "all_available" to keep all languages. For available language values, see https://www.wikidata.org/wiki/Help:Wikimedia_
id_df	Default to NULL. If given, it should be a dataframe typically generated with tw_get_(), and is used instead of calling Wikidata or using SQLite cache. Ignored when id is of length more than one.
cache	Defaults to NULL. If given, it should be given either TRUE or FALSE. Typically set with tw_enable_cache() or tw_disable_cache().
overwrite_cache	Logical, defaults to FALSE. If TRUE, it overwrites the table in the local sqlite database. Useful if the original Wikidata object has been updated.
cache_connection	Defaults to NULL. If NULL, and caching is enabled, tidywikidatar will use a local sqlite database. A custom connection to other databases can be given (see vignette caching for details).
disconnect_db	Defaults to TRUE. If FALSE, leaves the connection to cache open.
wait	In seconds, defaults to 0. Time to wait between queries to Wikidata. If data are cached locally, wait time is not applied. If you are running many queries systematically you may want to add some waiting time between queries.

Value

A list of the same length of input (or a character vector is only_first is set to TRUE)

Examples

```

# By default, it returns a list of the same length as input,
# no matter how many values for each id/property
tw_get_property_same_length(
  id = c(

```

```

      "Q180099",
      "Q228822",
      "Q76857"
    ),
    p = "P26"
  )

if (interactive()) {
  # Notice that if no relevant match is found, it returns a NA
  # This is useful for piped operations

  tibble::tibble(id = c(
    "Q180099",
    "Q228822",
    "Q76857"
  )) %>%
    dplyr::mutate(spouse = tw_get_property_same_length(id, "P26"))

  # Consider unnesting for further analysis

  tibble::tibble(id = c(
    "Q180099",
    "Q228822",
    "Q76857"
  )) %>%
    dplyr::mutate(spouse = tw_get_property_same_length(id, "P26")) %>%
    tidyr::unnest(cols = spouse)

  # If you are sure that you are interested only in the first return value,
  # consider setting only_first=TRUE to get a character vector rather than a list
  # Be mindful: you may well be discarding valid values.
  tibble::tibble(id = c(
    "Q180099",
    "Q228822",
    "Q76857"
  )) %>%
    dplyr::mutate(spouse = tw_get_property_same_length(id, "P26",
      only_first = TRUE
    ))
}

```

tw_get_property_with_details

Gets all details of a property

Description

Gets all details of a property

Usage

```
tw_get_property_with_details(id, p, wait = 0)
```

Arguments

id	A character vector, must start with Q, e.g. "Q254" for Wolfgang Amadeus Mozart.
p	A character vector, a property. Must always start with the capital letter "P", e.g. "P31" for "instance of".
wait	In seconds, defaults to 0. Time to wait between queries to Wikidata. If data are cached locally, wait time is not applied. If you are running many queries systematically you may want to add some waiting time between queries.

Value

A tibble, corresponding to the details for the given property. NULL if no relevant property found.

Examples

```
# Get "female form of label", including language
tw_get_property_with_details(id = "Q64733534", p = "P2521")
```

```
tw_get_property_with_details_single
Gets all details of a property
```

Description

Gets all details of a property

Usage

```
tw_get_property_with_details_single(id, p)
```

Arguments

id	A character vector, must start with Q, e.g. "Q254" for Wolfgang Amadeus Mozart.
p	A character vector, a property. Must always start with the capital letter "P", e.g. "P31" for "instance of".

Value

A tibble, corresponding to the details for the given property. NULL if no relevant property found.

Examples

```
# Get "female form of label", including language
tw_get_property_with_details(id = "Q64733534", p = "P2521")
```

tw_get_qualifiers	<i>Get Wikidata qualifiers for a given property of a given item</i>
-------------------	---

Description

N.B. In order to provide for consistently structured output, this function outputs either id or value for each qualifier. The user should keep in mind that some of these come with additional detail (e.g. the unit, precision, or reference calendar).

Usage

```
tw_get_qualifiers(
  id,
  p,
  language = tidywikidatar::tw_get_language(),
  cache = NULL,
  overwrite_cache = FALSE,
  cache_connection = NULL,
  disconnect_db = TRUE,
  wait = 0,
  include_id_and_p = TRUE
)
```

Arguments

id	A character vector of length 1, must start with Q, e.g. "Q254" for Wolfgang Amadeus Mozart.
p	A character vector of length 1, a property. Must always start with the capital letter "P", e.g. "P31" for "instance of".
language	Defaults to language set with <code>tw_set_language()</code> ; if not set, "en". Use "all_available" to keep all languages. For available language values, see https://www.wikidata.org/wiki/Help:Wikimedia
cache	Defaults to NULL. If given, it should be given either TRUE or FALSE. Typically set with <code>tw_enable_cache()</code> or <code>tw_disable_cache()</code> .
overwrite_cache	Logical, defaults to FALSE. If TRUE, it overwrites the table in the local sqlite database. Useful if the original Wikidata object has been updated.
cache_connection	Defaults to NULL. If NULL, and caching is enabled, <code>tidywikidatar</code> will use a local sqlite database. A custom connection to other databases can be given (see vignette caching for details).
disconnect_db	Defaults to TRUE. If FALSE, leaves the connection to cache open.
wait	In seconds, defaults to 0. Time to wait between queries to Wikidata. If data are cached locally, wait time is not applied. If you are running many queries systematically you may want to add some waiting time between queries.
include_id_and_p	Logical, defaults to TRUE. If TRUE, output includes a column with the wikidata id of the item.

Value

A data frame (a tibble) with five columns: `id` for the input id, `qualifier_id`, `property`, `value`, and `set` (to distinguish sets of data when a property is present more than once)

Examples

```
tw_get_qualifiers(id = "Q180099", p = "P26", language = "en")
```

```
tw_get_qualifiers_single
```

Get Wikidata qualifiers for a given property of a given item

Description

N.B. In order to provide for consistently structured output, this function outputs either `id` or `value` for each qualifier. The user should keep in mind that some of these come with additional detail (e.g. the unit, precision, or reference calendar).

Usage

```
tw_get_qualifiers_single(  
  id,  
  p,  
  language = tidywikidatar::tw_get_language(),  
  cache = NULL,  
  overwrite_cache = FALSE,  
  cache_connection = NULL,  
  disconnect_db = TRUE,  
  wait = 0,  
  include_id_and_p = TRUE  
)
```

Arguments

<code>id</code>	A character vector of length 1, must start with Q, e.g. "Q254" for Wolfgang Amadeus Mozart.
<code>p</code>	A character vector of length 1, a property. Must always start with the capital letter "P", e.g. "P31" for "instance of".
<code>language</code>	Defaults to language set with <code>tw_set_language()</code> ; if not set, "en". Use "all_available" to keep all languages. For available language values, see https://www.wikidata.org/wiki/Help:Wikimedia
<code>cache</code>	Defaults to NULL. If given, it should be given either TRUE or FALSE. Typically set with <code>tw_enable_cache()</code> or <code>tw_disable_cache()</code> .
<code>overwrite_cache</code>	Logical, defaults to FALSE. If TRUE, it overwrites the table in the local sqlite database. Useful if the original Wikidata object has been updated.

cache_connection	Defaults to NULL. If NULL, and caching is enabled, tidywikidatar will use a local sqlite database. A custom connection to other databases can be given (see vignette caching for details).
disconnect_db	Defaults to TRUE. If FALSE, leaves the connection to cache open.
wait	In seconds, defaults to 0. Time to wait between queries to Wikidata. If data are cached locally, wait time is not applied. If you are running many queries systematically you may want to add some waiting time between queries.
include_id_and_p	Logical, defaults to TRUE. If TRUE, output includes a column with the wikidata id of the item.

Value

A data frame (a tibble) with five columns: id for the input id, qualifier_id, property, value, and set (to distinguish sets of data when a property is present more than once)

Examples

```
tw_get_qualifiers(id = "Q180099", p = "P26", language = "en")
```

tw_get_single	<i>Return (most) information from a Wikidata item in a tidy format</i>
---------------	--

Description

Return (most) information from a Wikidata item in a tidy format

Usage

```
tw_get_single(
  id,
  language = tidywikidatar::tw_get_language(),
  cache = NULL,
  overwrite_cache = FALSE,
  read_cache = TRUE,
  cache_connection = NULL,
  disconnect_db = TRUE,
  wait = 0
)
```

Arguments

id	A character vector, must start with Q, e.g. "Q180099" for the anthropologist Margaret Mead. Can also be a data frame of one row, typically generated with tw_search() or a combination of tw_search() and tw_filter_first().
----	--

language	Defaults to language set with <code>tw_set_language()</code> ; if not set, "en". Use "all_available" to keep all languages. For available language values, see https://www.wikidata.org/wiki/Help:Wikimedia_
cache	Defaults to NULL. If given, it should be given either TRUE or FALSE. Typically set with <code>tw_enable_cache()</code> or <code>tw_disable_cache()</code> .
overwrite_cache	Logical, defaults to FALSE. If TRUE, it overwrites the table in the local sqlite database. Useful if the original Wikidata object has been updated.
read_cache	Logical, defaults to TRUE. Mostly used internally to prevent checking if an item is in cache if it is already known that it is not in cache.
cache_connection	Defaults to NULL. If NULL, and caching is enabled, tidywikidatar will use a local sqlite database. A custom connection to other databases can be given (see vignette caching for details).
disconnect_db	Defaults to TRUE. If FALSE, leaves the connection to cache open.
wait	In seconds, defaults to 0. Time to wait between queries to Wikidata. If data are cached locally, wait time is not applied. If you are running many queries systematically you may want to add some waiting time between queries.

Value

A data.frame (a tibble) with three columns (id, property, and value). If item not found or trouble connecting with the server, an data frame with three columns and zero rows is returned, with the warning as an attribute, which can be retrieved with `attr(output, "warning")`

Examples

```
tw_get_single(
  id = "Q180099",
  language = "en"
)
```

tw_get_wikipedia	<i>Get Wikipedia article in given language</i>
------------------	--

Description

Get Wikipedia article in given language

Usage

```
tw_get_wikipedia(
  id,
  language = tidywikidatar::tw_get_language(),
  id_df = NULL,
  cache = NULL,
  overwrite_cache = FALSE,
```



```

    cache_connection = NULL,
    disconnect_db = TRUE,
    wait = 0
)

```

Arguments

id	A character vector, must start with Q, e.g. "Q254" for Wolfgang Amadeus Mozart
language	Defaults to language set with <code>tw_set_language()</code> ; if not set, "en". Use "all_available" to keep all languages. For available language values, see https://www.wikidata.org/wiki/Help:Wikimedia
id_df	Default to NULL. If given, it should be a dataframe typically generated with <code>tw_get_()</code> , and is used instead of calling Wikidata or using SQLite cache. Ignored when id is of length more than one.
cache	Defaults to NULL. If given, it should be given either TRUE or FALSE. Typically set with <code>tw_enable_cache()</code> or <code>tw_disable_cache()</code> .
overwrite_cache	Logical, defaults to FALSE. If TRUE, it overwrites the table in the local sqlite database. Useful if the original Wikidata object has been updated.
cache_connection	Defaults to NULL. If NULL, and caching is enabled, <code>tidywikidatar</code> will use a local sqlite database. A custom connection to other databases can be given (see vignette caching for details).
disconnect_db	Defaults to TRUE. If FALSE, leaves the connection to cache open.
wait	In seconds, defaults to 0. Time to wait between queries to Wikidata. If data are cached locally, wait time is not applied. If you are running many queries systematically you may want to add some waiting time between queries.

Value

A character vector of length 1, with the Wikidata label in the requested language.

Examples

```
tw_get_wikipedia(id = "Q180099")
```

```
tw_get_wikipedia_base_api_url
```

Facilitates the creation of MediaWiki API base URLs

Description

Mostly used internally

Usage

```
tw_get_wikipedia_base_api_url(url = NULL, title = NULL, language = NULL)
```

Arguments

url	Full URL to a Wikipedia page. If given, title and language can be left empty.
title	Title of a Wikipedia page or final parts of its url. If given, url can be left empty, but language must be provided.
language	Two-letter language code used to define the Wikipedia version to use. If url given, this can be left empty.

Value

A character vector of base urls to be used with the MediaWiki API

Examples

```
tw_get_wikipedia_base_api_url(title = "Margaret Mead", language = "en")
```

tw_label	<i>Gets labels for all columns with names such as "id" and "property".</i>
----------	--

Description

Gets labels for all columns with names such as "id" and "property".

Usage

```
tw_label(
  df,
  value = TRUE,
  language = tidywikidatar::tw_get_language(),
  cache = NULL,
  overwrite_cache = FALSE,
  cache_connection = NULL,
  disconnect_db = TRUE,
  wait = 0
)
```

Arguments

df	A data frame, typically generated with other tidywikidatar functions such as tw_get_property()
value	Logical, defaults to TRUE. If TRUE, it tries to get labels for all supposed id in the column called value. May break if the columns include some value which starts with Q and some digits, but is not a wikidata id.
language	Defaults to language set with tw_set_language(); if not set, "en". Use "all_available" to keep all languages. For available language values, see https://www.wikidata.org/wiki/Help:Wikimedia_

cache	Defaults to NULL. If given, it should be given either TRUE or FALSE. Typically set with <code>tw_enable_cache()</code> or <code>tw_disable_cache()</code> .
overwrite_cache	Logical, defaults to FALSE. If TRUE, it overwrites the table in the local sqlite database. Useful if the original Wikidata object has been updated.
cache_connection	Defaults to NULL. If NULL, and caching is enabled, <code>tidywikidatar</code> will use a local sqlite database. A custom connection to other databases can be given (see vignette caching for details).
disconnect_db	Defaults to TRUE. If FALSE, leaves the connection to cache open.
wait	In seconds, defaults to 0. Time to wait between queries to Wikidata. If data are cached locally, wait time is not applied. If you are running many queries systematically you may want to add some waiting time between queries.

Value

A data frame, with the same shape as the input data frame, but with labels instead of identifiers.

Examples

```
tw_get_qualifiers(id = "Q180099", p = "P26", language = "en") %>%
  head(2) %>%
  tw_label()
```

tw_query	<i>Perform simple Wikidata queries</i>
----------	--

Description

This function aims to facilitate only the most basic type of queries: return which items have the following property pairs. For more details on Wikidata queries, consult: https://www.wikidata.org/wiki/Wikidata:SPARQL_queries. For complex queries, use `WikidataQueryServiceR::query_wikidata()`

Usage

```
tw_query(
  query,
  fields = c("item", "itemLabel", "itemDescription"),
  language = "en",
  return_as_tw_search = TRUE
)
```

Arguments

query	A list of named vectors, or a data frame (see example and readme).
fields	A character vector of Wikidata fields. Ignored if return_as_tw_search is set to TRUE (as per default). Defaults to ("item", "itemLabel", "itemDescription")
language	Two letter code of preferred language, defaults to "en". If more than one, can be set in order of preference, e.g. c("it", "fr", "en")
return_as_tw_search	Logical, defaults to TRUE. If TRUE, returns a data frame with three columns (id, label, and description) that can be piped to other tw_ functions. If FALSE, a data frame with as many columns as fields.

Value

A data frame

Examples

```
if (interactive()) {
  query <- list(
    c(p = "P106", q = "Q1397808"),
    c(p = "P21", q = "Q6581072")
  )
  tw_query(query)
}
```

tw_search	<i>Search for Wikidata items or properties and return Wikidata id, label, and description.</i>
-----------	--

Description

By defaults, this search returns items. Set type to property or use tw_search_property() for properties.

Usage

```
tw_search(
  search,
  type = "item",
  language = tidywikidatar::tw_get_language(),
  limit = 10,
  include_search = FALSE,
  wait = 0,
  cache = NULL,
  overwrite_cache = FALSE,
  cache_connection = NULL,
```

```

    disconnect_db = TRUE
  )

```

Arguments

search	A string to be searched in Wikidata
type	Defaults to "item". Either "item" or "property".
language	Language to be used for the search. Can be set once per session with <code>tw_set_language()</code> . If not set, defaults to "en". For a full list, see https://www.wikidata.org/wiki/Help:Wikimedia_language_codes .
limit	Maximum numbers of responses to be given.
include_search	Logical, defaults to FALSE. If TRUE, the search is returned as an additional column.
wait	In seconds, defaults to 0. Time to wait between queries to Wikidata. If data are cached locally, wait time is not applied. If you are running many queries systematically you may want to add some waiting time between queries.
cache	Defaults to NULL. If given, it should be given either TRUE or FALSE. Typically set with <code>tw_enable_cache()</code> or <code>tw_disable_cache()</code> .
overwrite_cache	Defaults to FALSE. If TRUE, overwrites cache.
cache_connection	Defaults to NULL. If NULL, and caching is enabled, <code>tidywikidatar</code> will use a local sqlite database. A custom connection to other databases can be given (see vignette caching for details).
disconnect_db	Defaults to TRUE. If FALSE, leaves the connection to cache open.

Value

A data frame (a tibble) with three columns (id, label, and description), and as many rows as there are results (by default, limited to 10). Four columns when `include_search` is set to TRUE.

Examples

```
tw_search(search = c("Margaret Mead", "Ruth Benedict"))
```

tw_search_item	<i>Search for Wikidata properties in Wikidata and return Wikidata id, label, and description.</i>
----------------	---

Description

This search returns only items, use `tw_search_property()` for properties.

Usage

```
tw_search_item(
  search,
  language = tidywikidatar::tw_get_language(),
  limit = 10,
  include_search = FALSE,
  wait = 0,
  cache = NULL,
  overwrite_cache = FALSE,
  cache_connection = cache_connection
)
```

Arguments

search	A string to be searched in Wikidata
language	Language to be used for the search. Can be set once per session with <code>tw_set_language()</code> . If not set, defaults to "en". For a full list, see https://www.wikidata.org/wiki/Help:Wikimedia_language_codes
limit	Maximum numbers of responses to be given.
include_search	Logical, defaults to FALSE. If TRUE, the search is returned as an additional column.
wait	In seconds, defaults to 0.1. Time to wait between queries to Wikidata. If data are cached locally, wait time is not applied.
cache	Defaults to NULL. If given, it should be given either TRUE or FALSE. Typically set with <code>tw_enable_cache()</code> or <code>tw_disable_cache()</code> .
overwrite_cache	Defaults to FALSE. If TRUE, overwrites cache.
cache_connection	Defaults to NULL. If NULL, and caching is enabled, <code>tidywikidatar</code> will use a local sqlite database. A custom connection to other databases can be given (see vignette caching for details).

Value

A data frame (a tibble) with three columns (id, label, and description), and as many rows as there are results (by default, limited to 10).

Examples

```
tw_search_item(search = "Sylvia Pankhurst")
```

tw_search_property	<i>Search for Wikidata properties in Wikidata and return Wikidata id, label, and description.</i>
--------------------	---

Description

This search returns only properties, use `tw_search_items()` for properties.

Usage

```
tw_search_property(
  search,
  language = tidywikidatar::tw_get_language(),
  limit = 10,
  include_search = FALSE,
  wait = 0,
  cache = NULL,
  overwrite_cache = FALSE,
  disconnect_db = TRUE,
  cache_connection = NULL
)
```

Arguments

search	A string to be searched in Wikidata.
language	Language to be used for the search. Can be set once per session with <code>tw_set_language()</code> . If not set, defaults to "en". For a full list, see https://www.wikidata.org/wiki/Help:Wikimedia_language_codes .
limit	Maximum numbers of responses to be given.
include_search	Logical, defaults to FALSE. If TRUE, the search is returned as an additional column.
wait	In seconds, defaults to 0.1. Time to wait between queries to Wikidata. If data are cached locally, wait time is not applied.
cache	Defaults to NULL. If given, it should be given either TRUE or FALSE. Typically set with <code>tw_enable_cache()</code> or <code>tw_disable_cache()</code> .
overwrite_cache	Defaults to FALSE. If TRUE, overwrites cache.
disconnect_db	Defaults to TRUE. If FALSE, leaves the connection to cache open.
cache_connection	Defaults to NULL. If NULL, and caching is enabled, <code>tidywikidatar</code> will use a local sqlite database. A custom connection to other databases can be given (see vignette caching for details).

Value

A data frame (a tibble) with three columns (id, label, and description), and as many rows as there are results (by default, limited to 10).

Examples

```
tw_search_property(search = "gender")
```

tw_search_single	<i>Search for Wikidata items or properties and return Wikidata id, label, and description.</i>
------------------	--

Description

This search returns only items, use `tw_search_property()` for properties.

Usage

```
tw_search_single(
  search,
  type = "item",
  language = tidywikidatar::tw_get_language(),
  limit = 10,
  include_search = FALSE,
  cache = NULL,
  overwrite_cache = FALSE,
  cache_connection = NULL,
  disconnect_db = TRUE,
  wait = 0
)
```

Arguments

search	A string to be searched in Wikidata
type	Defaults to "item". Either "item" or "property".
language	Language to be used for the search. Can be set once per session with <code>tw_set_language()</code> . If not set, defaults to "en". For a full list, see https://www.wikidata.org/wiki/Help:Wikimedia_language_codes
limit	Maximum numbers of responses to be given.
include_search	Logical, defaults to FALSE. If TRUE, the search is returned as an additional column.
cache	Defaults to NULL. If given, it should be given either TRUE or FALSE. Typically set with <code>tw_enable_cache()</code> or <code>tw_disable_cache()</code> .
overwrite_cache	Defaults to FALSE. If TRUE, overwrites cache.
cache_connection	Defaults to NULL. If NULL, and caching is enabled, <code>tidywikidatar</code> will use a local sqlite database. A custom connection to other databases can be given (see vignette caching for details).
disconnect_db	Defaults to TRUE. If FALSE, leaves the connection to cache open.
wait	In seconds, defaults to 0. Time to wait between queries to Wikidata. If data are cached locally, wait time is not applied. If you are running many queries systematically you may want to add some waiting time between queries.

Value

A data frame (a tibble) with three columns (id, label, and description), and as many rows as there are results (by default, limited to 10). Four columns when include_search is set to TRUE.

Examples

```
tw_search_single(search = "Sylvia Pankhurst")
```

```
tw_set_cache_folder    Set folder for caching data
```

Description

Consider using a folder out of your current project directory, e.g. `tw_set_cache_folder("~/R/tw_data/")`: you will be able to use the same cache in different projects, and prevent cached files from being sync-ed if you use services such as Nextcloud or Dropbox.

Usage

```
tw_set_cache_folder(path = NULL)
```

```
tw_get_cache_folder(path = NULL)
```

Arguments

path A path to a location used for caching data. If the folder does not exist, it will be created.

Value

The path to the caching folder, if previously set; the same path as given to the function; or the default, `tw_data` is none is given.

Examples

```
if (interactive()) {  
  tw_set_cache_folder(fs::path(fs::path_home_r(), "R", "tw_data"))  
}  
  
tw_get_cache_folder()
```

tw_set_language	<i>Set language to be used by all functions</i>
-----------------	---

Description

Defaults to "en".

Usage

```
tw_set_language(language = NULL)
```

```
tw_get_language(language = NULL)
```

Arguments

language	A character vector of length one, with a string of two letters such as "en". For a full list of available values, see: https://www.wikidata.org/wiki/Help:Wikimedia_language_codes/lists/all
----------	--

Value

A two letter code for the language, if previously set; the same language as given to the function; or the default, en is none is given.

Examples

```
if (interactive()) {  
  tw_set_language(language = "en")  
}
```

```
tw_get_language()
```

tw_write_item_to_cache	<i>Writes item to cache</i>
------------------------	-----------------------------

Description

Writes item to cache. Typically used internally, but exported to enable custom caching solutions.

Usage

```
tw_write_item_to_cache(
  item_df,
  language = tidywikidatar::tw_get_language(),
  overwrite_cache = FALSE,
  cache_connection = NULL,
  disconnect_db = TRUE
)
```

Arguments

`item_df` A data frame with three columns typically generated with `tw_get()`.

`language` Defaults to language set with `tw_set_language()`; if not set, "en". Use "all_available" to keep all languages. For available language values, see https://www.wikidata.org/wiki/Help:Wikimedia_

`overwrite_cache` Logical, defaults to FALSE. If TRUE, it first deletes all rows associated with the item(s) included in `item_df`. Useful if the original Wikidata object has been updated.

`cache_connection` Defaults to NULL. If NULL, and caching is enabled, `tidywikidatar` will use a local sqlite database. A custom connection to other databases can be given (see vignette caching for details).

`disconnect_db` Defaults to TRUE. If FALSE, leaves the connection to cache open.

Value

Nothing, used for its side effects.

Examples

```
tw_set_cache_folder(path = fs::path(tempdir(), paste(sample(letters, 24), collapse = "")))
tw_create_cache_folder(ask = FALSE)
tw_disable_cache()

df_from_api <- tw_get(id = "Q180099", language = "en")

df_from_cache <- tw_get_cached_item(
  id = "Q180099",
  language = "en"
)

is.null(df_from_cache) # expect TRUE, as nothing has yet been stored in cache

tw_write_item_to_cache(
  item_df = df_from_api,
  language = "en"
)
```

```
df_from_cache <- tw_get_cached_item(
  id = "Q180099",
  language = "en"
)

is.null(df_from_cache) # expect a data frame, same as df_from_api
```

```
tw_write_qualifiers_to_cache
  Write qualifiers to cache
```

Description

Mostly to be used internally by tidywikidatar, use with caution to keep caching consistent.

Usage

```
tw_write_qualifiers_to_cache(
  qualifiers_df,
  language = tidywikidatar::tw_get_language(),
  overwrite_cache = FALSE,
  cache_connection = NULL,
  disconnect_db = TRUE
)
```

Arguments

qualifiers_df	A data frame with two columns typically generated with <code>tw_get(include_id_and_p = FALSE)</code> .
language	Defaults to language set with <code>tw_set_language()</code> ; if not set, "en". Use "all_available" to keep all languages. For available language values, see https://www.wikidata.org/wiki/Help:Wikimedia_
overwrite_cache	Logical, defaults to FALSE. If TRUE, it overwrites the table in the local sqlite database. Useful if the original Wikidata object has been updated.
cache_connection	Defaults to NULL. If NULL, and caching is enabled, tidywikidatar will use a local sqlite database. A custom connection to other databases can be given (see vignette caching for details).
disconnect_db	Defaults to TRUE. If FALSE, leaves the connection to cache open.

Value

Silently returns the same data frame provided as input. Mostly used internally for its side effects.

Examples

```

q_df <- tw_get_qualifiers(
  id = "Q180099",
  p = "P26",
  language = "en",
  cache = FALSE
)

tw_write_qualifiers_to_cache(
  qualifiers_df = q_df,
  language = "en"
)

```

```

tw_write_search_to_cache
Writes search to cache

```

Description

Writes search to cache. Typically used internally, but exported to enable custom caching solutions.

Usage

```

tw_write_search_to_cache(
  search_df,
  type = "item",
  language = tidywikidatar::tw_get_language(),
  overwrite_cache = FALSE,
  cache_connection = NULL,
  disconnect_db = TRUE
)

```

Arguments

search_df	A data frame with four columns typically generated with <code>tw_search(include_search = TRUE)</code> .
type	Defaults to "item". Either "item" or "property".
language	Defaults to language set with <code>tw_set_language()</code> ; if not set, "en". Use "all_available" to keep all languages. For available language values, see https://www.wikidata.org/wiki/Help:Wikimedia_
overwrite_cache	Logical, defaults to FALSE. If TRUE, it first deletes all rows associated with the item(s) included in the provided search_df.
cache_connection	Defaults to NULL. If NULL, and caching is enabled, tidywikidatar will use a local sqlite database. A custom connection to other databases can be given (see vignette caching for details).
disconnect_db	Defaults to TRUE. If FALSE, leaves the connection to cache open.

Value

Nothing, used for its side effects.

Examples

```
tw_set_cache_folder(path = fs::path(tempdir(), paste(sample(letters, 24), collapse = "")))
tw_create_cache_folder(ask = FALSE)
tw_disable_cache()

search_from_api <- tw_search(search = "Sylvia Pankhurst", include_search = TRUE)

search_from_cache <- tw_get_cached_search("Sylvia Pankhurst")

nrow(search_from_cache) == 0 # expect TRUE, as nothing has yet been stored in cache

tw_write_search_to_cache(search_df = search_from_api)

search_from_cache <- tw_get_cached_search("Sylvia Pankhurst")

search_from_cache
```

Index

tw_check_cache, 3
tw_check_cache_folder, 4
tw_check_cached_items, 3
tw_check_qid, 5
tw_check_search, 6
tw_connect_to_cache, 7
tw_create_cache_folder, 8
tw_disable_cache, 8
tw_disconnect_from_cache, 9
tw_enable_cache, 10
tw_filter, 10
tw_filter_first, 11
tw_filter_people, 12
tw_get, 13
tw_get_cache_file, 18
tw_get_cache_folder
 (tw_set_cache_folder), 49
tw_get_cache_table_name, 18
tw_get_cached_item, 14
tw_get_cached_qualifiers, 15
tw_get_cached_search, 17
tw_get_description, 19
tw_get_field, 20
tw_get_id_of_wikipedia_page, 21
tw_get_image, 22
tw_get_image_metadata, 23
tw_get_image_metadata_single, 24
tw_get_image_same_length, 26
tw_get_label, 27
tw_get_language (tw_set_language), 50
tw_get_links_from_wikipedia_page, 29
tw_get_property, 30
tw_get_property_description, 31
tw_get_property_label, 32
tw_get_property_same_length, 33
tw_get_property_with_details, 35
tw_get_property_with_details_single,
 36
tw_get_qualifiers, 37
tw_get_qualifiers_single, 38
tw_get_single, 39
tw_get_wikipedia, 40
tw_get_wikipedia_base_api_url, 41
tw_label, 42
tw_query, 43
tw_search, 44
tw_search_item, 45
tw_search_property, 47
tw_search_single, 48
tw_set_cache_folder, 49
tw_set_language, 50
tw_write_item_to_cache, 50
tw_write_qualifiers_to_cache, 52
tw_write_search_to_cache, 53