

Package: scythe (via r-universe)

October 25, 2024

Title Harvest and register data package citations

Version 1.1.0

Description Harvests data package citations from several API sources, including PLOS, Scopus, and Springer. This package uses modified functions from `rplos`, which is no longer maintained.

License Apache License (>= 2.0)

Encoding UTF-8

LazyData true

Imports curl, dplyr, jsonlite, keyring, rcrossref, solrium, stats

Remotes ropensci/bib2df@a8e96e13f5

Suggests bib2df, covr, purrr, testthat (>= 3.0.0)

RoxygenNote 7.3.1

Config/testthat/edition 3

Repository <https://dataoneorg.r-universe.dev>

RemoteUrl <https://github.com/DataONEorg/scythe>

RemoteRef HEAD

RemoteSha 73c3c5849bd6d20a24bceeb2405bb7541a388b4d

Contents

citation_search	2
citation_search_plos	2
citation_search_scopus	3
citation_search_springer	4
citation_search_xdd	4
ploscompact	5
report_est_wait	5
scythe_get_key	6
scythe_set_key	6
searchplos	7
strextact	8
write_citation_pairs	8

Index**9**

citation_search	<i>Search for citations in text across all APIs</i>
-----------------	---

Description

Search for citations in text across all APIs

Usage

```
citation_search(identifiers, sources = c("plos", "scopus", "springer", "xdd"))
```

Arguments

`identifiers` a vector of identifiers to be searched for
`sources` a vector indicating which sources to query (one or more of plos, scopus, springer)

Value

tibble of matching dataset and publication identifiers

Examples

```
## Not run:  
identifiers <- c("10.18739/A22274", "10.18739/A2D08X", "10.5063/F1T151VR")  
result <- citation_search(identifiers, sources = c("plos"))  
  
## End(Not run)
```

citation_search_plos	<i>Search for citations in PLOS</i>
----------------------	-------------------------------------

Description

This function searches for citations in PLOS. Requests are throttled at one identifier every 6 seconds so as to not overload the PLOS API. This function uses modified source code from the ‘rplos’ package, which is no longer maintained.

Usage

```
citation_search_plos(identifiers)
```

Arguments

`identifiers` a vector of identifiers to be searched for

Value

tibble of matching dataset and publication identifiers

Examples

```
## Not run:  
identifiers <- c("10.18739/A22274", "10.18739/A2D08X", "10.5063/F1T151VR")  
result <- citation_search_plos(identifiers)  
  
## End(Not run)
```

citation_search_scopus

Search for citations in Scopus

Description

This function searches for citations in Scopus. Requests are throttled 9 requests/second so as to not overload the Scopus API.

Usage

```
citation_search_scopus(identifiers)
```

Arguments

`identifiers` a vector of identifiers to be searched for

Value

tibble of matching dataset and publication identifiers

Examples

```
## Not run:  
identifiers <- c("10.18739/A22274", "10.18739/A2D08X", "10.5063/F1T151VR")  
result <- citation_search_scopus(identifiers)  
  
## End(Not run)
```

`citation_search_springer` *Search for citations from Springer*

Description

This function searches for citations from Springer. It requires that an API key be obtained from [Springer](https://dev.springernature.com/) and set using `'scythe_set_key()'`. Requests are throttled at one identifier every second so as to not overload the PLOS API.

Usage

```
citation_search_springer(identifiers)
```

Arguments

`identifiers` a vector of identifiers to be searched for

Value

tibble of matching dataset and publication identifiers

Examples

```
## Not run:
identifiers <- c("10.18739/A22274", "10.18739/A2D08X", "10.5063/F1T151VR")
result <- citation_search_springer(identifiers)

## End(Not run)
```

`citation_search_xdd` *Search for citations in xDD*

Description

This function searches for citations in xDD. Uses `'snippets/term'` function in xDD API and searches through all of xDD corpus (not limited to full-text documents).

Usage

```
citation_search_xdd(identifiers)
```

Arguments

`identifiers` a vector of identifiers to be searched for, without hypertext transfer protocol: `"https://"` or `"http://"`

Value

tibble of publications and their identifiers that contain

Examples

```
## Not run:
identifiers <- c("10.18739/A22274", "10.18739/A2D08X", "10.5063/F1T151VR", "10.18739/A29K97")
result <- citation_search_xdd(identifiers)

## End(Not run)
```

ploscompact	<i>This function is from the 'rplos' package, which is no longer maintained.</i>
-------------	--

Description

This function is from the 'rplos' package, which is no longer maintained.

Usage

```
ploscompact(l)
```

Arguments

l	a list
---	--------

report_est_wait	<i>Report estimated wait for rate limited queries</i>
-----------------	---

Description

Report estimated wait for rate limited queries

Usage

```
report_est_wait(n_queries, wait_seconds)
```

Arguments

n_queries	number of queries
wait_seconds	wait time in seconds between queries

scythe_get_key *Get API secret for a key source.*

Description

Look for API keys for services, which are represented as character strings.

Usage

```
scythe_get_key(source)
```

Arguments

source the name of the source service to look up

Details

Secrets are typically stored in a keyring named "scythe" (see the keyring package), or, alternatively, in an environment variable with a name identical to "source".

Value

character the secret value, or NA if not set

Examples

```
## Not run:  
scythe_get_key("scopus_key")  
  
## End(Not run)
```

scythe_set_key *Set API keys for search services*

Description

This function sets API keys using the 'keyring' package. 'keyring' uses your operating system's credential store to securely keep track of key-value pairs. Running this function for the first time will prompt you to set a password for your keyring, should you need to lock or unlock it. See 'keyring::keyring_unlock' for more details.

Usage

```
scythe_set_key(source, secret)
```

Arguments

source	(char) Key source, one of "scopus" or "springer"
secret	(char) API key value

searchplos	<i>A Modified Version of rplos::searchplos</i>
------------	--

Description

This function is adapted from the searchplos in the 'rplos' package, which is no longer maintained.

Usage

```
searchplos(
  q = NULL,
  fl = "id",
  fq = NULL,
  sort = NULL,
  start = 0,
  limit = 10,
  sleep = 6,
  errors = "simple",
  proxy = NULL,
  callopts = list(),
  progress = NULL,
  ...
)
```

Arguments

q	Search terms, eg: field:query
fl	Fields to return
fq	Fields to filter query on
sort	Sort results according to field
start	Record to start at for pagination
limit	Number of results to return for pagination
sleep	Seconds to wait between requests
errors	One of simple or complete
proxy	List of args for proxy connection
callopts	Optional curl options
progress	Optional logic for progress bar
...	Addtl Solr arguments

streextract	<i>This function is from the 'rplos' package, which is no longer maintained.</i>
-------------	--

Description

This function is from the 'rplos' package, which is no longer maintained.

Usage

```
streextract(str, pattern)
```

Arguments

str	A string
pattern	A regex pattern

write_citation_pairs	<i>Write citation pairs</i>
----------------------	-----------------------------

Description

Write citation pairs

Usage

```
write_citation_pairs(citation_list, path, extra_fields = NULL)
```

Arguments

citation_list	(data.frame) data.frame of citation pairs containing variables article_id and dataset_id
path	(char) path to write JSON citation pairs to
extra_fields	(char) list of extra fields to pass to bib2df::bib2df

Examples

```
## Not run:  
pairs <- data.frame(article_id = "10.1371/journal.pone.0213037",  
                   dataset_id = "10.18739/A22274")  
write_citation_pairs(citation_list = pairs, path = "citation_pairs.json")  
  
## End(Not run)
```


Index

`citation_search`, 2
`citation_search_plos`, 2
`citation_search_scopus`, 3
`citation_search_springer`, 4
`citation_search_xdd`, 4

`ploscompact`, 5

`report_est_wait`, 5

`scythe_get_key`, 6
`scythe_set_key`, 6
`searchplos`, 7
`strextract`, 8

`write_citation_pairs`, 8