

Package ‘rdataretriever’

March 11, 2017

Title R Interface to the Data Retriever

Description Provides an R interface to the Data Retriever <<http://data-retriever.org/>> via the Data Retriever's command line interface. The Data Retriever automates the tasks of finding, downloading, and cleaning public datasets, and then stores them in a local database.

Version 1.0.0

Date 2017-02-21

BugReports <https://github.com/ropensci/rdataretriever/issues>

URL <https://github.com/ropensci/rdataretriever/>

Depends R (>= 3.0.0)

Imports utils

SystemRequirements Data Retriever (>= 1.8)

License MIT + file LICENSE

LazyData true

RoxygenNote 5.0.1

NeedsCompilation no

Author Daniel McGlinn [aut, cre],
Henry Senyondo [aut],
Shawn Taylor [aut],
Ethan White [aut]

Maintainer Daniel McGlinn <danmcglinn@gmail.com>

Repository CRAN

Date/Publication 2017-03-11 09:18:44

R topics documented:

datasets	2
download	2
fetch	3

get_updates	3
install	4
rdataretriever	5
reset	5

Index	6
--------------	----------

datasets	<i>Name all available dataset scripts.</i>
----------	--

Description

Additional information on the available datasets can be found at <https://retriever.readthedocs.io/en/latest/datasets.html>

Usage

```
datasets()
```

Value

returns a character vector with the available datasets for download

Examples

```
rdataretriever::datasets()
```

download	<i>Download datasets via the Data Retriever.</i>
----------	--

Description

Directly downloads data files with no processing, allowing downloading of non-tabular data.

Usage

```
download(dataset, path = ".", sub_dir = FALSE, log_dir = NULL)
```

Arguments

dataset	the name of the dataset that you wish to download
path	the path where the data should be downloaded to
sub_dir	if true and the downloaded dataset is stored in subdirectories those subdirectories will be preserved and placed according the path argument, defaults to false.
log_dir	the location where the retriever log should be stored if the progress is not printed to the console

Examples

```
rdataretriever::download('portal')
## list files downloaded
dir('.', pattern='portal')
```

fetch

Fetch a dataset via the Data Retriever

Description

Each datafile in a given dataset is downloaded to a temporary directory and then imported as a `data.frame` as a member of a named list.

Usage

```
fetch(dataset, quiet = TRUE)
```

Arguments

dataset	the name of the dataset that you wish to download
quiet	logical, if true retriever runs in quiet mode

Examples

```
## fetch the portal Database
portal = rdataretriever::fetch('portal')
class(portal)
names(portal)
## preview the data in the portal species datafile
head(portal$species)
```

get_updates

Update the retriever's dataset scripts to the most recent versions.

Description

This function will check if the version of the retriever's scripts in your local directory `~/retriever/scripts/` is up-to-date with the most recent official retriever release. Note it is possible that even more updated scripts exist at the retriever repository <https://github.com/weecology/retriever/tree/master/scripts> that have not yet been incorporated into an official release, and you should consider checking that page if you have any concerns.

Usage

```
get_updates()
```

Examples

```
rdataretriever::get_updates()
```

```
install
```

```
Install datasets via the Data Retriever.
```

Description

Data is stored in either CSV files or one of the following database management systems: MySQL, PostgreSQL, SQLite, or Microsoft Access.

Usage

```
install(dataset, connection, db_file = NULL, conn_file = NULL,
        data_dir = ".", log_dir = NULL)
```

Arguments

dataset	the name of the dataset that you wish to download
connection	what type of database connection should be used. The options include: mysql, postgres, sqlite, msaccess, or csv'
db_file	the name of the database file the dataset should be loaded into
conn_file	the path to the .conn file that contains the connection configuration options for mysql and postgres databases. This defaults to mysql.conn or postgres.conn respectively. The connection file is a file that is formatted in the following way:
	<pre> host my_server@my_host.com port my_port_number user my_user_name password my_password </pre>
data_dir	the location where the dataset should be installed. Only relevant for csv connection types. Defaults to current working directory
log_dir	the location where the retriever log should be stored if the progress is not printed to the console

Examples

```
rdataretriever::install('iris', 'csv')
```

rdataretriever	<i>rdataretriever: R interface to the Data Retriever.</i>
----------------	---

Description

rdataretriever: R interface to the Data Retriever.

About

The Data Retriever automates the tasks of finding, downloading, and cleaning up publicly available data, and then stores them in a local database or flat files. This lets data analysts spend less time cleaning up and managing data, and more time analyzing it.

This package lets you access the Data Retriever using R, so that the Retriever's data handling can easily be integrated into R workflows.

reset	<i>Reset rdataretriever.</i>
-------	------------------------------

Description

Reset the components of rdataretriever using scope [all, scripts, data, connection]

Usage

```
reset(scope = "all")
```

Arguments

scope	what components to reset. Options include: 'scripts', 'data', 'connection' and 'all', where 'all' is the default setting that resets all components.
-------	--

Examples

```
rdataretriever::reset()
```

Index

*Topic **utilities**

get_updates, 3

datasets, 2

download, 2

fetch, 3

get_updates, 3

install, 4

rdataretriever, 5

rdataretriever-package
(rdataretriever), 5

reset, 5