

Package ‘taxizedb’

June 21, 2017

Type Package

Title Tools for Working with 'Taxonomic' Databases

Description Tools for working with 'taxonomic' databases, including utilities for downloading databases, loading them into various 'SQL' databases, cleaning up files, and providing a 'SQL' connection that can be used to do 'SQL' queries directly or used in 'dplyr'.

Version 0.1.4

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

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

License MIT + file LICENSE

LazyData TRUE

Imports curl (>= 2.4), DBI (>= 0.6-1), RPostgreSQL (>= 0.4.1), RMySQL (>= 0.10.11), RSQLite (>= 1.1.2), dplyr (>= 0.7.0), dbplyr (>= 1.0.0), magrittr (>= 1.5), hoardr (>= 0.1.0)

Suggests roxygen2 (>= 6.0.1), testthat

RoxygenNote 6.0.1

NeedsCompilation no

Author Scott Chamberlain [aut, cre]

Maintainer Scott Chamberlain <myrmecocystus+r@gmail.com>

Repository CRAN

Date/Publication 2017-06-20 23:54:44 UTC

R topics documented:

taxizedb-package	2
db_download	3
db_load	4
sql_collect	6
src_taxizedb	7
tdb_cache	8

Index	9
--------------	----------

taxizedb-package	<i>Taxonomic databases interface</i>
------------------	--------------------------------------

Description

Taxonomic databases interface

Database setup/user/pwd

Every user will have a unique combination of username, password, and platform, so we can't make the functions in this package work out of the box in every situation.

Author(s)

Scott Chamberlain <myrmecocystus@gmail.com>

Examples

```
## Not run:
# IMPORTANT: Remember to start your PostgreSQL database for ITIS
# and ThePlantList and your MySQL database for COL

# data source: ITIS
## download ITIS database
x <- db_download_itis()
db_load_itis(x)

## connect to the ITIS database
src <- src_itis()

## use SQL syntax
sql_collect(src, "select * from hierarchy limit 5")
### or pipe the src to sql_collect
src %>% sql_collect("select * from hierarchy limit 5")

## use dplyr verbs
src %>%
  tbl("hierarchy") %>%
  top_n(10)

## or create tbl object for repeated use
hiers <- src %>% tbl("hierarchy")
hiers %>% top_n(10)
hiers %>% select(tsn, level)

# data source: theplantlist
## download tpl datababase
x <- db_download_tpl()
db_load_tpl(x)
```

```
## connecto the tpl database
src <- src_tpl()

## do queries
tpl <- src %>% tbl("plantlist")
tpl %>% filter(family == "Pinaceae")

# data source: catalogue of life
## download col datababase
x <- db_download_col()
db_load_col(x)

## connecto the col database
src <- src_col()

## do queries
cnames <- src %>% tbl("common_name_element")
cnames %>% select(name)

## End(Not run)
```

db_download

Download taxonomic databases

Description

Download taxonomic databases

Usage

```
db_download_itis(verbose = TRUE)
```

```
db_download_tpl(verbose = TRUE)
```

```
db_download_col(verbose = TRUE)
```

```
db_download_gbif(verbose = TRUE)
```

Arguments

verbose (logical) Print messages. Default: TRUE

Details

Downloads sql database, cleans up unneeded files, returns path to sql file

Value

Path to the downloaded SQL database

Supported

- ITIS - PostgreSQL
- The PlantList - PostgreSQL
- Catalogue of Life - MySQL
- GBIF - SQLite

Beware

COL database loading takes a long time, e.g., 30 minutes. you may want to run it in a separate R session, or just look at the db_load_col fxn and run the commands in your shell.

See Also

[tdb_cache](#)

Examples

```
## Not run:
# ITIS
# x <- db_download_itis()
# db_load_itis(x)
# src_itis()

# Plantlist
# x <- db_download_tpl()
# db_load_tpl(x, "sacmac")
# src_tpl()

# COL
# x <- db_download_col()
# db_load_col(x)
# src_col()

# GBIF
# x <- db_download_gbif()
# db_load_gbif()
# src_gbif(x)

## End(Not run)
```

db_load

Load taxonomic databases

Description

Load taxonomic databases

Usage

```
db_load_itis(path, user, pwd = NULL, verbose = TRUE)

db_load_tpl(path, user, pwd = NULL, verbose = TRUE)

db_load_col(path, user = "root", pwd = NULL, verbose = TRUE)

db_load_gbif(verbose = TRUE)
```

Arguments

path	(character) path to the .sql database file
user	(character) User name
pwd	(character) Password, if any
verbose	(logical) Print messages. Default: TRUE

Details

These functions load SQL DBs into the respective driver, and they return the file path, but they don't load the database

We check if the database used for each source is installed on your machine first. and if it is running or not, with errors for the user if neither are true.

Value

Nothing, just message on success

Supported

- ITIS - PostgreSQL
- the PlantList - PostgreSQL
- Catalogue of Life - MySQL
- GBIF - SQLite

Beware

COL database loading takes a long time, e.g., 30 minutes. you may want to run it in a separate R session, or just look at the db_load_col fxn and run the commands in your shell.

Examples

```
## Not run:
# ITIS
# x <- db_download_itis()
# db_load_itis(x, "<your user name>", "<your password>")

# Plantlist
# x <- db_download_tpl()
```

```
# db_load_tpl(x, "<your user name>", "<your password>")

# COL
x <- db_download_col()
db_load_col(x, "<your user name>", "<your password>")

# GBIF
## only checks if sqlite installed
db_load_gbif()

## End(Not run)
```

 sql_collect

Query and get data back into a data.frame

Description

Query and get data back into a data.frame

Usage

```
sql_collect(src, query, ...)
```

Arguments

src	An src object, result of calling <code>src_itis()</code> , <code>src_col()</code> , or <code>src_tpl()</code>
query	A SQL query
...	further args passed on to <code>dplyr::tbl()</code>

Details

we run `dplyr::tbl()`, then `dplyr::collect()`

Examples

```
## Not run:
src <- src_itis()
sql_collect(src, "select * from hierarchy limit 5")
## or pipe the src to sql_collect
src %>% sql_collect("select * from hierarchy limit 5")

## End(Not run)
```

src_taxizedb	<i>src - dplyr src objects</i>
--------------	--------------------------------

Description

src - dplyr src objects

Usage

```
src_itis(user, password, dbname = "ITIS", ...)
```

```
src_tpl(user, password, dbname = "plantlist", ...)
```

```
src_col(user = "root", password = NULL, dbname = "col", ...)
```

```
src_gbif(path)
```

Arguments

user (character) user name

password (character) password

dbname (character) database name. Defaults: ITIS, col, and plantlistdb for ITIS, COL, and ThePlantlist, respectively. GBIF uses SQLite so doesn't have a database name

... Further args passed on to [DBI::dbConnect\(\)](#)

path (character) path to SQLite database

Value

an src object

Examples

```
## Not run:  
# src_itis()  
# src_tpl()  
# src_col()  
# src_gbif()  
  
## End(Not run)
```

tdb_cache

Caching

Description

Manage cached taxizedb files with **hoardr**

Details

cache_delete only accepts 1 file name, while cache_delete_all doesn't accept any names, but deletes all files. For deleting many specific files, use cache_delete in a `lapply()` type call

Useful user functions

- `tdb_cache$cache_path_get()` get cache path
- `tdb_cache$cache_path_set()` set cache path
- `tdb_cache$list()` returns a character vector of full path file names
- `tdb_cache$files()` returns file objects with metadata
- `tdb_cache$details()` returns files with details
- `tdb_cache$delete()` delete specific files
- `tdb_cache$delete_all()` delete all files, returns nothing

Examples

```
## Not run:
tdb_cache

# list files in cache
tdb_cache$list()

# delete certain database files
# tdb_cache$delete("file path")
# tdb_cache$list()

# delete all files in cache
# tdb_cache$delete_all()
# tdb_cache$list()

## End(Not run)
```


Index

*Topic **package**

taxizedb-package, 2

db_download, 3

db_download_col (db_download), 3

db_download_gbif (db_download), 3

db_download_itis (db_download), 3

db_download_tpl (db_download), 3

db_load, 4

db_load_col (db_load), 4

db_load_gbif (db_load), 4

db_load_itis (db_load), 4

db_load_tpl (db_load), 4

DBI::dbConnect(), 7

dplyr::collect(), 6

dplyr::tbl(), 6

lapply(), 8

sql_collect, 6

src_col (src_taxizedb), 7

src_col(), 6

src_gbif (src_taxizedb), 7

src_itis (src_taxizedb), 7

src_itis(), 6

src_taxizedb, 7

src_tpl (src_taxizedb), 7

src_tpl(), 6

taxizedb (taxizedb-package), 2

taxizedb-package, 2

tdb_cache, 4, 8