Package 'pubtatordb'

October 14, 2022

Type Package

Title Create and Query a Local 'PubTator' Database

Version 0.1.4

Maintainer Zachary Colburn <zcolburn@gmail.com>

Description

'PubTator' <https://www.ncbi.nlm.nih.gov/CBBresearch/Lu/Demo/PubTator/> is a National Center for Biotechnology Information (NCBI) tool that enhances the annotation of articles on PubMed <https://www.ncbi.nlm.nih.gov/pubmed/>. It makes it possible to rapidly identify potential relationships between genes or proteins using text mining techniques. In contrast, manually searching for and reading the annotated articles would be very time consuming. 'PubTator' offers both an online interface and a RESTful API, however, neither of these approaches are well suited for frequent, high-throughput analyses. The package 'pubtatordb' provides a set of functions that make it easy for the average R user to download 'PubTator' annotations, create, and then query a local version of the database.

License MIT + file LICENSE

Encoding UTF-8

LazyData true

Suggests covr, testthat, knitr, rmarkdown

Imports DBI, R.utils, RSQLite, assertthat, dplyr, readr

VignetteBuilder knitr

RoxygenNote 7.0.0

NeedsCompilation no

Author Zachary Colburn [aut, cre],
Madigan Army Medical Center - Department of Clinical Investigation
[cph, fnd]

Repository CRAN

Date/Publication 2019-11-22 19:30:02 UTC

R topics documented:

---

download_pt   *Download PubTator data via ftp.*

Description

Download PubTator data via ftp.

Usage

```
download_pt(pubtator_parent_path, ...)
```

Arguments

pubtator_parent_path

> The path to the directory where the PubTator data folder will be created.

...   Additional arguments to dir.create and download.file.

Value

The path to the newly created directory. This can be passed to other functions as the pt_path argument.

Examples

```
# Use the full path. The files are large. Writing somewhere other than the
# temp directory is recommended.
download_path <- tempdir()
download_pt(dowload_path)
```

```
make_pubtator_sqlite_path
```
                    Make a path to the PubTator sqlite file.

Description

Make a path to the PubTator sqlite file.

Usage

```
make_pubtator_sqlite_path(pt_path)
```

Arguments

pt_path   A character string indicating the full path of the directory containing the pubtator
                 gz files to be extracted.

Value

A character string indicating the full path to the sqlite file.

```
pt_columns                  List the column names for a table in the PubTator sqlite database
```

Description

List the column names for a table in the PubTator sqlite database

Usage

```
pt_columns(db_con, table_name)
```

Arguments

db_con   A connection to the PubTator sqlite database, as created via pubator_connector.
table_name   The name of the table of interest. Valid tables can be found using pt_tables.
                 Capitalization does not matter.

Value

A character vector of the column names for a given table.

Examples

```
db_con <- pt_connector(pt_path)
pubtator_columns(db_con, "gene")
```

---

pt_connector   *Connect to pubtator.sqlite*

Description

Connect to pubtator.sqlite

Usage

```
pt_connector(pt_path)
```

Arguments

pt_path   A character string indicating the full path of the directory containing the pubtator
                 gz files to be extracted.

Value

A SQLiteConnection

Examples

```
pt_connector("D:/Reference_data/PubTator")
```

---

pt_select   *Retrieve data from the PubTator database.*

Description

Retrieve data from the PubTator database.

Usage

```
pt_select(
  db_con,
  table_name,
  columns = NULL,
  keys = NULL,
  keytype = NULL,
  limit = Inf
)
```

Arguments

`db_con`	A connection to the PubTator sqlite database, as created via pubator_connector.
`table_name`	The name of the table of interest. Valid tables can be found using pt_tables. Capitalization does not matter.
`columns`	A character vector of the names of the columns of interest. Capitalization does not matter.
`keys`	A vector specifying which values must be in the keytype column to enable retrieval. No filtering is performed if keys = NULL.
`keytype`	The column in which the keys should be searched for.
`limit`	The maximum number of rows the query should return. All rows passing filtering (if any) are returned if limit = Inf.

Value

A data.frame.

Examples

```
db_con <- pt_connector(pt_path)
pt_select(
  db_con,
  "gene",
  columns = c("ENTREZID","Resource","MENTIONS","PMID"),
  keys = c("7356", "4199", "7018"),
  keytype = "ENTREZID",
  limit = 10
)
```

---

pt_tables	*List the tables in the PubTator sqlite database*

Description

List the tables in the PubTator sqlite database

Usage

```
pt_tables(db_con)
```

Arguments

`db_con`	A connection to the PubTator sqlite database, as created via pubator_connector.

Value

A character vector of the names of the tables found in the database.

Examples

```
db_con <- pt_connector(pt_path)
pt_tables(db_con)
```

---

pt_to_sql   *Create sqlite database from the pubtator data.*

Description

Create sqlite database from the pubtator data.

Usage

```
pt_to_sql(pt_path, skip_behavior = TRUE, remove_behavior = FALSE)
```

Arguments

pt_path	A character string indicating the full path of the directory containing the pubtator gz files to be extracted.
skip_behavior	TRUE/FALSE indicating whether the file should be re-extracted if it has already been extracted.
remove_behavior	
	TRUE/FALSE indicating whether the gz files should be removed following successful extraction.

Examples

```
download_path <- tempdir()
current_dir <- getwd()
setwd(download_path)
pt_to_sql("PubTator")
setwd(current_dir)
```

---

pubtator_citations   *See the citations for PubTator*

Description

See the citations for PubTator

Usage

```
pubtator_citations()
```

Examples

```
pubtator_citations()
```

---

pubtator_ftp_url   *NCBI's ftp url definition for PubTator.*

Description

NCBI's ftp url definition for PubTator.

Usage

```
pubtator_ftp_url()
```

Value

A character string giving the ftp url for PubTator.

---

pubtator_tables   *Table and dataset definitions*

Description

Table and dataset definitions

Usage

```
pubtator_tables()
```

Value

A character vector where names are table names and values are dataset names.

Index