Package 'GEOmetadb'

April 16, 2019

Type Package

Title A compilation of metadata from NCBI GEO

Version 1.44.0

Date 2017-11-22

Depends GEOquery,RSQLite

Suggests knitr, rmarkdown, dplyr, tm, wordcloud

Author Jack Zhu and Sean Davis

Maintainer Jack Zhu <zhujack@mail.nih.gov>

biocViews Infrastructure

Description The NCBI Gene Expression Omnibus (GEO) represents the largest public repository of microarray data. However, finding data of interest can be challenging using current tools. GEOmetadb is an attempt to make access to the metadata associated with samples, platforms, and datasets much more feasible. This is accomplished by parsing all the NCBI GEO metadata into a SQLite database that can be stored and queried locally. GEOmetadb is simply a thin wrapper around the SQLite database along with associated documentation. Finally, the SQLite database is updated regularly as new data is added to GEO and can be downloaded at will for the most up-to-date metadata. GEOmetadb paper: http://bioinformatics.oxfordjournals.org/cgi/content/short/24/23/2798 .

URL http://gbnci.abcc.ncifcrf.gov/geo/

VignetteBuilder knitr

License Artistic-2.0

git_url https://git.bioconductor.org/packages/GEOmetadb

git_branch RELEASE_3_8

git_last_commit 72a4529

git_last_commit_date 2018-10-30

Date/Publication 2019-04-15

R topics documented:

GEOmetadb-package *Query NCBI GEO metadata from a local SQLite database*

Description

The NCBI Gene Expression Omnibus (GEO) represents the largest public repository of microarray data. However, finding data of interest can be challenging using current tools. GEOmetadb is an attempt to make access to the metadata associated with samples, platforms, and datasets much more feasible. This is accomplished by parsing all the NCBI GEO metadata into a SQLite database that can be stored and queried locally. GEOmetadb is simply a thin wrapper around the SQLite database along with associated documentation. Finally, the SQLite database is updated regularly as new data is added to GEO and can be downloaded at will for the most up-to-date metadata.

Details

Package:	GEOmetadb
Type:	Package
Version:	1.1.5
Date:	2008-09-09
License:	Artistic-2.0

Author(s)

Jack Zhu and Sean Davis

Maintainer: Jack Zhu <zhujack@mail.nih.gov>

References

http://meltzerlab.nci.nih.gov/apps/geo, http://gbnci.abcc.ncifcrf.gov/geo/

Examples

```
if(file.exists('GEOmetadb.sqlite')) {
  a <- columnDescriptions()[1:5,]
  b <- geoConvert('GPL97','GSM')
} else {
  print("use getSQLiteFile() to get a copy of the GEOmetadb SQLite file
and then rerun the example")
}
```

columnDescriptions *Get column descriptions for the GEOmetadb database*

Description

Searching the GEOmetadb database requires a bit of knowledge about the structure of the database and column descriptions. This function returns those column descriptions for all columns in all tables in the database.

Usage

```
columnDescriptions(sqlite_db_name='GEOmetadb.sqlite')
```

Arguments

sqlite_db_name	The filename of the GEOmetadb sqlite database file

Value

A three-column data.frame including TableName, FieldName, and Description.

Author(s)

Sean Davis <sdavis2@mail.nih.gov>

References

http://meltzerlab.nci.nih.gov/apps/geo

Examples

```
if(file.exists('GEOmetadb.sqlite')) {
  columnDescriptions()[1:5,]
} else {
  print("You will need to usethe getSQLiteFile() function to get a copy
of the SQLite database file before this example will work")
}
```

---

geoConvert *Cross-reference between GEO data types*

Description

A common task is to find all the GEO entities of one type associated with another GEO entity (eg., find all GEO samples associated with GEO platform 'GPL96'). This function provides a very fast mapping between entity types to facilitate queries of this type.

Usage

```
geoConvert(in_list, out_type = c("gse", "gpl", "gsm", "gds", "smatrix"), sqlite_db_name = "GEOmeta
```

Arguments

in_list	Character vector of GEO entities to convert from.
out_type	Character vector of GEO entity types to which to convert.
sqlite_db_name	The filename of the GEOmetadb sqlite database file

Value

A list of data.frames.

Author(s)

Jack Zhu <zhujack@mail.nih.gov>

References

http://meltzerlab.nci.nih.gov/apps/geo, http://gbnci.abcc.ncifcrf.gov/geo/

Examples

```
if(file.exists("GEOmetadb.sqlite")) {
  geoConvert('GPL96',out_type='GSM')
} else {
  print("Run getSQLiteFile() to get a copy of the GEOmetadb SQLite file
and then rerun the example")
}
```

---

getBiocPlatformMap *Get mappings between GPL and Bioconductor microarry annotation packages*

Description

Query the gpl table and get GPL information of a given list of Bioconductor microarry annotation packages. Note currently the GEOmetadb does not contains all the mappings, but we are trying to construct a relative complete list.

Usage

```
getBiocPlatformMap(con, bioc='all')
```

Arguments

con	Connection to the GEOmetadb.sqlite database
bioc	Character vector of Biocondoctor microarry annotation packages, e.g. c('hgu133plus2','hgu95av2'). 'all' returns all mappings.

Value

A six-column data.frame including GPL title, GPL accession, bioc_package, manufacturer, organism, data_row_count.

Author(s)

Jack Zhu <zhujack@mail.nih.gov>, Sean Davis <sdavis2@mail.nih.gov>

References

http://meltzerlab.nci.nih.gov/apps/geo

Examples

```
if(file.exists('GEOmetadb.sqlite')) {
    con <- dbConnect(SQLite(), "GEOmetadb.sqlite")
getBiocPlatformMap(con)[1:5,]
getBiocPlatformMap(con, bioc=c('hgu133a','hgu95av2'))
dbDisconnect(con)
} else {
  print("You will need to usethe getSQLiteFile() function to get a copy
of the SQLite database file before this example will work")
}
```

`getSQLiteFile` *Download and unzip the most recent GEOmetadb SQLite file*

Description

This function is the standard method for downloading and unzipping the most recent GEOmetadb SQLite file from the server.

Usage

```
getSQLiteFile(destdir = getwd(), destfile = "GEOmetadb.sqlite.gz")
```

Arguments

`destdir`	The destination directory of the downloaded file
`destfile`	The filename of the downloaded file. This filename should end in ".gz" as the unzipping assumes that is the case

Value

Prints some diagnostic information to the screen.

Returns the local filename for use later.

Author(s)

Sean Davis <sdavis2@mail.nih.gov>

References

http://meltzerlab.nci.nih.gov/apps/geo, http://gbnci.abcc.ncifcrf.gov/geo/

Examples

```
## Not run: geometadbfile <- getSQLiteFile()
```

Index