

# Package ‘GEOsearch’

November 3, 2016

**Type** Package

**Title** GEOsearch

**Version** 1.0.0

**Date** 2015-07-10

**Author** Zhicheng Ji, Hongkai Ji

**Maintainer** Zhicheng Ji <zji4@jhu.edu>

**Description** GEOsearch is an extendable search engine for NCBI GEO (Gene Expression Omnibus). Instead of directly searching the term, GEOsearch can find all the gene names contained in the search term and search all the alias of the gene names simultaneously in GEO database. GEOsearch also provides other functions such as summarizing common biology keywords in the search results.

**License** GPL(>=2)

**Imports** RCurl, org.Hs.eg.db, org.Mm.eg.db

**VignetteBuilder** knitr

**Suggests** knitr, shiny, DT, org.Ag.eg.db, org.At.tair.db, org.Bt.eg.db, org.Ce.eg.db, org.Cf.eg.db, org.Dm.eg.db, org.Dr.eg.db, org.EcK12.eg.db, org.EcSakai.eg.db, org.Gg.eg.db, org.Mmu.eg.db, org.Pf.plasmo.db, org.Pt.eg.db, org.Rn.eg.db, org.Sc.sgd.db, org.Ss.eg.db, org.Xl.eg.db

**Depends** R(>= 3.2)

**biocViews** GUI,Software

**RoxygenNote** 5.0.1

**NeedsCompilation** no

## R topics documented:

BatchDownload . . . . .	2
database . . . . .	2
GEOsearch . . . . .	3
GEOsearchTerm . . . . .	3
GEOsearchUI . . . . .	4
KeyWordFreq . . . . .	4
SampleDetail . . . . .	5
term . . . . .	6
TermAlias . . . . .	6
<b>Index</b>	<b>8</b>

---

BatchDownload	<i>BatchDownload</i>
---------------	----------------------

---

### Description

Generate shell file to download the SRA of multiple GSE or GSM

### Usage

```
BatchDownload(namelist, GSMdownloadpath = ".")
```

### Arguments

namelist	A character list of GSE or GSM ID.
GSMdownloadpath	A character value specifying the path to download the SRA files on users own Linux/Unix computer.

### Details

This function returns a shell file that can download the SRA of multiple GSE or GSM. The shell file should be run on user's own computer with Linux/Unix system and wget installed. The download process is to set up a bunch of directories under the path specified by users (GSMdownloadpath) and download the SRA files inside the paths. Note: it is the users' job to make sure the GSE and GSM names are correct. Otherwise the function will run forever due to inner error control system. Also, it could take a long time for the function to complete if there are lots of GSE series in namelist.

### Value

A character string of commands to be run on users' computer.

### Author(s)

Zhicheng Ji, Hongkai Ji <zji4@zji4.edu>

### Examples

```
BatchDownload("GSM1274542")
```

---

database	<i>List of species database</i>
----------	---------------------------------

---

### Description

The list of species annotation databases used in alias searching. All databases are from Bioconductor.

### Format

A data.frame with two columns. First column: name of Bioconductor database; Second column: corresponding species name.

**Source**

<http://www.bioconductor.org/>

---

GEOsearch

*GEOsearch: Extendable Search Engine for Gene Expression Omnibus*

---

**Description**

GEOsearch provides extended search function (i.e. search all aliases of a gene name) as well as other utilities.

**Details**

GEOsearch is an extendable search engine for NCBI GEO (Gene Expression Omnibus). Instead of directly searching the term, GEOsearch can find all the gene names contained in the search term and search all the alias of the gene names simultaneously in GEO database. GEOsearch also provides other functions such as summarizing common biology keywords in the search results.

---

GEOsearchTerm

*GEOsearchTerm*

---

**Description**

Perform searching in NCBI GEO database.

**Usage**

```
GEOsearchTerm(termlist, type = "GSE")
```

**Arguments**

termlist	A character vector of terms to be searched in NCBI GEO. Typically the direct output of function "termaliases"
type	Either "GSE" or "GSM". The function will search either GSE and GSM on GEO.

**Details**

Search the terms one by one in NCBI GEO database and return an integrated table of search results. The returned results should contain exactly the same information as the results returned by directly searching in <http://www.ncbi.nlm.nih.gov/geo/>.

**Value**

A data frame containing the search results returned from NCBI GEO.

**Author(s)**

Zhicheng Ji, Hongkai Ji <zji4@zji4.edu>

**Examples**

```
GEOsearchTerm("Oct4 RNA-seq")
```

---

GEOsearchUI

*GEOsearchUI*

---

**Description**

Launch the GEOsearch graphical user interface

**Usage**

```
GEOsearchUI()
```

**Details**

This function will automatically launch the GEOsearch user interface in a web browser. The user interface provides many powerful functions (e.g. second-round search) which is not available by command line programming. It also provides a much easier and more convenient way to perform all functions available in command lines. The user interface can also be accessed by <http://zhiji.shinyapps.io/GEOsearch>. Neither R nor any packages are required in this online version. However, it is highly recommended that the user interface be launched locally for faster running speed.

**Author(s)**

Zhicheng Ji, Hongkai Ji <zji4@zji4.edu>

**Examples**

```
if(interactive()) {
  GEOsearchUI()
}
```

---

KeyWordFreq

*KeyWordFreq*

---

**Description**

Frequencies of common biology keywords appearing in search results

**Usage**

```
KeyWordFreq(searchtable, category = c("celltype", "disease", "tissue"))
```

**Arguments**

**searchtable** The direct output of function "GEOsearchterm"  
**category** A character vector specifying which category in the common biology keyword list to be used. Should be contain "celltype", "disease" or "tissue".

**Details**

This function calculates the frequencies of each common biology keyword appearing in the given search table. The list of common biology keywords is compiled from <http://www.atcc.org/>. The list contains three categories: cell types, diseases and tissues. Users can specify which category to be used. The function also returns log fold change and FDR of fisher test to check whether each keyword has significantly more appearance compared to base frequency. The base frequency is defined as the number of appearance of the key word in all samples (roughly 40000 samples) included in GEO database.

**Value**

A data.frame with the frequency of each common biology keyword. First column: keyword name; Second column: keyword frequency; Third column: log fold change of the frequency; Fourth column: FDR of fisher test.

**Author(s)**

Zhicheng Ji, Hongkai Ji <zji4@zji4.edu>

**Examples**

```
KeywordFreq(GEOSearchTerm("Oct4 RNA-seq"))
```

---

SampleDetail

*SampleDetail*

---

**Description**

Details of GSM samples given GSE accession ID.

**Usage**

```
SampleDetail(GSEid)
```

**Arguments**

GSEid                    A character vector of GSE accession ID.

**Details**

This function returns an integrated table containing details of all GSM samples for a list of GSE accession ID.

**Value**

A data frame containing the search results returned from NCBI GEO. First column: GEO Series Accession Number; Second column: GEO Sample Accession Number; Third column: Title; Fourth column: Type of experiment; Fifth column: Source; Sixth column: Organism; Seventh column: Characteristic; Eighth column: Description.

**Author(s)**

Zhicheng Ji, Hongkai Ji <zji4@zji4.edu>

**Examples**

```
SampleDetail(c("GSE69322", "GSE64008"))
```

---

term	<i>List of common biology keywords</i>
------	--

---

**Description**

The list of common biology keywords is compiled from <http://www.atcc.org/>. The list contains three categories: cell types, diseases and tissues.

**Format**

A data.frame with three columns. First column: keyword name; Second column: category; Third column: number of appearance in GEO database.

**Source**

<http://www.atcc.org/>

---

TermAlias	<i>TermAlias</i>
-----------	------------------

---

**Description**

Search Alias of a Term

**Usage**

```
TermAlias(searchterm, allspecies = c("Human", "Mouse"), mincount = 5)
```

**Arguments**

searchterm	A character value specifying the term being searched. Words should be separated by space.
allspecies	A character vector specifying the species in which the alias of gene names will be searched for. Available species: Anopheles, Arabidopsis, Bovine, Worm, Canine, Fly, Zebrafish, E coli strain K12, E coli strain Sakai, Chicken, Human, Mouse, Rhesus, Malaria, Chimp, Rat, Yeast, Pig, Xenopus.
mincount	An integer value specifying the minimum number of appearance in the GEO database for the alias to be retained. If an alias has too few appearance it may not be of great interest.

**Details**

This function first picks gene names from the search term. It then searches alias for all the gene names contained in the search term. It next queries GEO and retain alias that appear frequently enough in GEO database. Finally it returns a combinatory results of retained alias.

**Value**

A character vector of combinatory results of alias.

**Author(s)**

Zhicheng Ji, Hongkai Ji <zji4@zji4.edu>

**Examples**

```
TermAlias("Oct4 RNA-seq")
```

# Index

BatchDownload, [2](#)

database, [2](#)

GEOsearch, [3](#)

GEOsearch-package (GEOsearch), [3](#)

GEOsearchTerm, [3](#)

GEOsearchUI, [4](#)

KeywordFreq, [4](#)

SampleDetail, [5](#)

term, [6](#)

TermAlias, [6](#)