

Package ‘ontoProc’

October 16, 2018

Title processing of ontologies of anatomy, cell lines, and so on

Description Support harvesting of diverse bioinformatic ontologies, making particular use of the ontologyIndex package on CRAN. We provide snapshots of key ontologies for terms about cells, cell lines, chemical compounds, and anatomy, to help analyze genome-scale experiments, particularly cell x compound screens. Another purpose is to strengthen development of compelling use cases for richer interfaces to emerging ontologies.

Version 1.2.1

Author Vince Carey <stvjc@channing.harvard.edu>

Imports Biobase, S4Vectors, methods, AnnotationDbi, stats, utils, shiny

Suggests knitr, org.Hs.eg.db, org.Mm.eg.db, testthat

Depends R (>= 3.4), ontologyIndex

Maintainer VJ Carey <stvjc@channing.harvard.edu>

License Artistic-2.0

LazyLoad yes

biocViews Infrastructure, GO

RoxygenNote 6.0.1.9000

VignetteBuilder knitr

git_url <https://git.bioconductor.org/packages/ontoProc>

git_branch RELEASE_3_7

git_last_commit 4162d1e

git_last_commit_date 2018-05-11

Date/Publication 2018-10-15

R topics documented:

allGOTerms	2
c,TermSet-method	2
cellTypeToGO	3
cleanCLOnames	4
demoApp	4
dropStop	5

fastGrep	5
getCellOnto	6
humrna	7
makeSelectInput	7
minicorpus	8
nomenCheckup	9
secLevGen	9
siblings_TAG	10
stopWords	11
TermSet-class	11

Index 13

allGOterms	<i>allGOterms: data.frame with ids and terms</i>
------------	--

Description

allGOterms: data.frame with ids and terms

Usage

```
allGOterms
```

Format

data.frame instance

Source

This is a snapshot of all the terms available from GO.db (3.4.2), August 2017, using keys(GO.db, keytype="TERM").

Examples

```
data(allGOterms)
head(allGOterms)
```

c,TermSet-method	<i>combine TermSet instances</i>
------------------	----------------------------------

Description

combine TermSet instances

Usage

```
## S4 method for signature 'TermSet'
c(x, ...)
```

Arguments

x TermSet instance
 ... additional instances

Value

TermSet instance

cellTypeToGO	<i>utilities for approximate matching of cell type terms to GO categories and annotations</i>
--------------	---

Description

utilities for approximate matching of cell type terms to GO categories and annotations

Usage

```
cellTypeToGO(celltypeString, gotab, ...)

cellTypeToGenes(celltypeString, gotab, orgDb, cols = c("ENSEMBL", "SYMBOL"),
  ...)
```

Arguments

celltypeString character atom to be used to search GO terms using
 gotab a data.frame with columns GO (goids) and TERM (term strings) [agrep](#)
 ... additional arguments to [agrep](#)
 orgDb instances of orgDb
 cols columns to be retrieved in select operation

Value

data.frame
 data.frame

Note

Very primitive, uses [agrep](#) to try to find relevant terms.

Examples

```
data(allGOterms)
library(org.Hs.eg.db)
head(cellTypeToGO("serotonergic neuron", allGOterms))
head(cellTypeToGenes("serotonergic neuron", allGOterms, org.Hs.eg.db))
```

`cleanCLOnames` *obtain named character vector of terms from Cell Line Ontology, omitting obsolete and trailing 'cell'*

Description

obtain named character vector of terms from Cell Line Ontology, omitting obsolete and trailing 'cell'

Usage

```
cleanCLOnames()
```

Examples

```
cleanCLOnames()[1:10]
```

`demoApp` *demonstrate the use of makeSelectInput*

Description

demonstrate the use of makeSelectInput

Usage

```
demoApp()
```

Value

Run only for side effect of starting a shiny app.

Examples

```
if (interactive()) {  
  require(shiny)  
  print(demoApp())  
}
```

 dropStop

dropStop is a utility for removing certain words from text data

Description

dropStop is a utility for removing certain words from text data

Usage

```
dropStop(x, drop, lower = TRUE, splitby = " ")
```

Arguments

x	character vector of strings to be cleaned
drop	character vector of words to scrub
lower	logical, if TRUE, x converted with tolower
splitby	character, used with strsplit to tokenize x

Value

a list with one element per input string, split by " ", with elements in drop removed

Examples

```
data(minicorpus)
minicorpus[1:3]
dropStop(minicorpus)[1:3]
```

 fastGrep

some fields of interest are lists, and grep per se should not be used – this function checks and uses grep within vapply when appropriate

Description

some fields of interest are lists, and grep per se should not be used – this function checks and uses grep within vapply when appropriate

Usage

```
fastGrep(patt, onto, field, ...)
```

Arguments

patt	a regular expression whose presence in field should be checked
onto	an ontologyIndex instance
field	the ontologyIndex component to be searched
...	passed to grep

Examples

```
co = getCellOnto()
co
clo = getCellLineOnto()
length(clo$id)
che = getChebiLite()
length(che$id)
efo = getEFOnto()
length(efo$id)
```

humrna

humrna: a data.frame of SRA metadata related to RNA-seq in humans

Description

humrna: a data.frame of SRA metadata related to RNA-seq in humans

Usage

```
humrna
```

Format

```
data.frame
```

Note

arbitrarily chosen from RNA-seq studies for taxon 9606

Source

NCBI SRA

Examples

```
data(humrna)
names(humrna)
head(humrna[, 1:5])
```

makeSelectInput

generate a selectInput control for an ontologyIndex slice

Description

generate a selectInput control for an ontologyIndex slice

Usage

```
makeSelectInput(onto, term, type = "siblings", inputId, label,
  multiple = TRUE, ...)
```

Arguments

onto	ontologyIndex instance
term	character(1) term used as basis for term list option set in the control
type	character(1) 'siblings' or 'children', relationship to 'term' that the options will satisfy
inputId	character(1) for use in server
label	character(1) for labeling in ui
multiple	logical(1) passed to selectInput
...	additional parameters passed to selectInput

Value

a [selectInput](#) control

Examples

```
makeSelectInput
```

minicorpus	<i>minicorpus: a vector of annotation strings found in 'study title' of SRA metadata.</i>
------------	---

Description

minicorpus: a vector of annotation strings found in 'study title' of SRA metadata.

Usage

```
minicorpus
```

Format

character vector

Note

arbitrarily chosen from titles of RNA-seq studies for taxon 9606

Source

NCBI SRA

Examples

```
data(minicorpus)
head(minicorpus)
```

nomenCheckup	<i>repair nomenclature mismatches (to curated term set) in a vector of terms</i>
--------------	--

Description

repair nomenclature mismatches (to curated term set) in a vector of terms

Usage

```
nomenCheckup(cand, namedOffic, n = 1, tagcolname = "tag", ...)
```

Arguments

cand	character vector of candidate terms
namedOffic	named character vector of curated terms, the names are regarded as tags, intended to be identifiers in curated ontologies
n	numeric(1) number of nearest neighbors to return
tagcolname	character(1) prefix used to name columns for tags in output
...	passed to adist

Value

a data.frame instance with 2n+1 columns (column 1 is candidate, remaining n pairs of columns are (term, tag) for n nearest neighbors as measured by [adist](#)).

Examples

```
candidates = c("JHH7", "HUT102", "HS739T", "NCIH716")
# the candidates are cell line names returned in the text dump from
# https://portals.broadinstitute.org/ccle/page?gene=AHR
# note that one must travel to the third nearest neighbor
# to find the match (and tag) for Hs 739.T
# in this example, we compare to cell line names in Cell Line Ontology
nomenCheckup(candidates, cleanCLOnames(), n=3, tagcolname="clo")
```

secLevGen	<i>simple generation of children of 'choices' given as terms, returned as TermSet</i>
-----------	---

Description

simple generation of children of 'choices' given as terms, returned as TermSet

Usage

```
secLevGen(choices, ont)
```

Arguments

choices vector of terms
 ont instance of ontology_index (S3) from ontologyIndex package

Value

TermSet instance

Examples

```
efoOnto = getEFOnto()
secLevGen( "disease", efoOnto )
```

siblings_TAG	<i>generate a TermSet with siblings of a given term, excluding that term by default</i>
--------------	---

Description

generate a TermSet with siblings of a given term, excluding that term by default
 acquire the label of an ontology subject tag
 acquire the labels of children of an ontology subject tag

Usage

```
siblings_TAG(Tagstring = "EFO:1001209", ontology, justSibs = TRUE)
label_TAG(Tagstring = "EFO:0000311", ontology)
children_TAG(Tagstring = "EFO:1001209", ontology)
```

Arguments

Tagstring a character(1) that identifies a term
 ontology instance of ontology_index (S3) from ontologyIndex
 justSibs character(1)

Value

TermSet instance
 character(1)
 TermSet instance

Note

for label_TAG, Tagstring may be a vector

Examples

```

efoOnto = getEF00nto()
siblings_TAG( "EFO:1001209", efoOnto )
efoOnto = getEF00nto()
label_TAG( "EFO:0000311", efoOnto )
efoOnto = getEF00nto()
children_TAG( ontology = efoOnto )

```

stopWords

stopWords: vector of stop words from xpo6.com

Description

stopWords: vector of stop words from xpo6.com

Usage

```
stopWords
```

Format

character vector

Note

"Stop words" are english words that are assumed to contribute limited semantic value in the analysis of free text.

Source

<http://xpo6.com/list-of-english-stop-words/>

Examples

```

data(stopWords)
head(stopWords)

```

TermSet-class

manage ontological data with tags and a DataFrame instance

Description

manage ontological data with tags and a DataFrame instance

abbreviated display for TermSet instances

Usage

```

## S4 method for signature 'TermSet'
show(object)

```

Arguments

object instance of TermSet class

Value

instance of TermSet

Examples

```
efoOnto = getEF0Onto()  
defsibs = siblings_TAG("EFO:1001209", efoOnto)  
class(defsibs)  
defsibs
```

Index

*Topic **datasets**

- allGOterms, 2
- humrna, 7
- minicorpus, 8
- stopWords, 11

adist, 9
agrep, 3
allGOterms, 2

c, TermSet-method, 2
cellTypeToGenes (cellTypeToGO), 3
cellTypeToGO, 3
children_TAG (siblings_TAG), 10
cleanCLOnames, 4

demoApp, 4
dropStop, 5

fastGrep, 5

getCellLineOnto (getCellOnto), 6
getCellOnto, 6
getCellosaurusOnto (getCellOnto), 6
getChebiLite (getCellOnto), 6
getChebiOnto (getCellOnto), 6
getEFOnto (getCellOnto), 6
getOncotreeOnto (getCellOnto), 6
getUBERON_NE (getCellOnto), 6

humrna, 7

label_TAG (siblings_TAG), 10

makeSelectInput, 7
minicorpus, 8

nomenCheckup, 9

secLevGen, 9
selectInput, 8
show (TermSet-class), 11
show, TermSet-method (TermSet-class), 11
siblings_TAG, 10
stopWords, 11

TermSet-class, 11
tolower, 5