Package ‘DOSE’

January 13, 2016

Type Package

Title Disease Ontology Semantic and Enrichment analysis

Version 2.8.2

Author Guangchuang Yu <guangchuangyu@gmail.com> with contributions from
Li-Gen Wang, Vladislav Petyuk and Giovanni Dall’Olio.

Maintainer Guangchuang Yu <guangchuangyu@gmail.com>

Description This package implements five methods proposed by
Resnik, Schlicker, Jiang, Lin and Wang respectively
for measuring semantic similarities among DO terms and
gene products. Enrichment analyses including hypergeometric
model and gene set enrichment analysis are also implemented
for discovering disease associations of high-throughput
biological data.

Depends R (>= 3.1.0)

Imports methods, plyr, qvalue, stats4, AnnotationDbi, DO.db, igraph,
scales, reshape2, graphics, GOSemSim, grid, ggplot2

Suggests org.Hs.eg.db, clusterProfiler, knitr, BiocStyle

VignetteBuilder knitr

License Artistic-2.0

URL https://github.com/GuangchuangYu/DOSE

BugReports https://github.com/GuangchuangYu/DOSE/issues

biocViews Annotation, Visualization, MultipleComparison,
GeneSetEnrichment, Pathways, Software

NeedsCompilation no

R topics documented:

  DOSE-package ......................................................... 3
  ALLEXTID ............................................................. 3
  barplot.enrichResult .............................................. 4
<table>
<thead>
<tr>
<th>R topics documented:</th>
</tr>
</thead>
<tbody>
<tr>
<td>clusterSim</td>
</tr>
<tr>
<td>cnetplot</td>
</tr>
<tr>
<td>cnetplot_internal</td>
</tr>
<tr>
<td>computeIC</td>
</tr>
<tr>
<td>DataSet</td>
</tr>
<tr>
<td>doSim</td>
</tr>
<tr>
<td>dotplot</td>
</tr>
<tr>
<td>enrich.internal</td>
</tr>
<tr>
<td>enrichDO</td>
</tr>
<tr>
<td>enrichMap</td>
</tr>
<tr>
<td>enrichNCG</td>
</tr>
<tr>
<td>enrichResult-class</td>
</tr>
<tr>
<td>EXTID2NAME</td>
</tr>
<tr>
<td>EXTID2TERMID</td>
</tr>
<tr>
<td>fortify.enrichResult</td>
</tr>
<tr>
<td>fortify.gseaResult</td>
</tr>
<tr>
<td>gene2DO</td>
</tr>
<tr>
<td>geneSim</td>
</tr>
<tr>
<td>getALLEG</td>
</tr>
<tr>
<td>getGeneSet</td>
</tr>
<tr>
<td>gsea</td>
</tr>
<tr>
<td>gseAnalyzer</td>
</tr>
<tr>
<td>gseaplot</td>
</tr>
<tr>
<td>gseaResult-class</td>
</tr>
<tr>
<td>list2graph</td>
</tr>
<tr>
<td>mclusterSim</td>
</tr>
<tr>
<td>netplot</td>
</tr>
<tr>
<td>plot</td>
</tr>
<tr>
<td>rebuildAnnoData</td>
</tr>
<tr>
<td>scaleNodeColor</td>
</tr>
<tr>
<td>setReadable</td>
</tr>
<tr>
<td>setting_graph.attributes</td>
</tr>
<tr>
<td>show</td>
</tr>
<tr>
<td>simplot</td>
</tr>
<tr>
<td>summary</td>
</tr>
<tr>
<td>TERM2NAME</td>
</tr>
<tr>
<td>TERMID2EXTID</td>
</tr>
<tr>
<td>theme_dose</td>
</tr>
</tbody>
</table>

Index | 29
Disease Ontology Semantic and Enrichment analysis Implemented five methods proposed by Resnik, Schlicker, Jiang, Lin and Wang respectively for measuring DO semantic similarities, and hypergeometric test for enrichment analysis.

Description

This package is designed to estimate DO-based semantic similarity measurement and enrichment analysis.

Details

Package: DOSE
Type: Package
Version: 2.3.5
Date: 2-27-2012
biocViews: Bioinformatics, Annotation
Depends:
Imports: methods, AnnotationDbi, DO.db
Suggests: clusterProfiler, GOSemSim
License: Artistic-2.0

Author(s)

Guangchuang Yu, Li-Gen Wang
Maintainer: Guangchuang Yu <guangchuangyu@gmail.com>

See Also

enrichResult

Description

Get all background External ID.

Usage

ALLEXTID(organism, ...)

Description

Get all background External ID.
**Arguments**

organism organism
...

**Description**

barplot

**Usage**

```r
## S3 method for class 'enrichResult'
barplot(height, font.size = 12, title = "", ...)
```

**Arguments**

- height: enrichResult object
- font.size: font size
- title: plot title
- ...: other parameter, ignored

---

**clusterSim**

**Description**

semantic similarity between two gene clusters

**Usage**

```r
clusterSim(cluster1, cluster2, measure = "Wang", combine = "BMA")
```

**Arguments**

- cluster1: a vector of gene IDs
- cluster2: another vector of gene IDs
- measure: One of "Resnik", "Lin", "Rel", "Jiang" and "Wang" methods.
- combine: One of "max", "average", "rcmax", "BMA" methods, for combining

**Details**

given two gene clusters, this function calculates semantic similarity between them.


## Value

similarity

## Author(s)

Yu Guangchuang

## Examples

```r
## cluster1 <- c("835", "5261", "241", "994")
## clusterSim(cluster1, cluster2, ont="MF", organism="human", measure="Wang")
```

## Description

cnetplot method

## Usage

```r
cnetplot(x, showCategory = 5, categorySize = "geneNum", foldChange = NULL, fixed = TRUE, ...)
```

```
## S4 method for signature 'enrichResult'
cnetplot(x, showCategory = 5,
          categorySize = "geneNum", foldChange = NULL, fixed = TRUE, ...)
```

## Arguments

- `x` : enrichResult object
- `showCategory` : number of category plotted
- `categorySize` : one of geneNum or pvalue
- `foldChange` : fold change of expression value
- `fixed` : logical
- `...` : additional parameters

## Value

plot

## Author(s)

Guangchuang Yu [http://ygc.name](http://ygc.name)
cnetplot_internal

description
plot function of gene concept net.

usage
cnetplot_internal(inputlist, categorysize = "geneNum", showCategory = 5,
pvalue = NULL, foldChange = NULL, fixed = TRUE, DE.foldChange = NULL,
...
)

arguments
inputList a list of gene IDs
categorySize setting category size
showCategory number of categories to plot
pvalue pvalue
foldChange fold Change
fixed logical
DE.foldChange logical
... additional parameters

value
plotted igraph object.

author(s)
Guangchuang Yu http://ygc.name

computeIC

description
compute information content

usage
computeIC(ont = "DO", organism = "human")
Arguments

<table>
<thead>
<tr>
<th>Argument</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>ont</td>
<td>&quot;DO&quot;</td>
</tr>
<tr>
<td>organism</td>
<td>&quot;human&quot;</td>
</tr>
</tbody>
</table>

Author(s)

Guangchuang Yu [http://ygc.name](http://ygc.name)

---

**Description**

Information content and DO term to entrez gene IDs mapping

**doSim**

doSIm

**Description**

measuring similarities between two DO term vectors.

**Usage**

doSIm(DOID1, DOID2, measure = "Wang")

**Arguments**

<table>
<thead>
<tr>
<th>Argument</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>DOID1</td>
<td>DO term vector</td>
</tr>
<tr>
<td>DOID2</td>
<td>DO term vector</td>
</tr>
<tr>
<td>measure</td>
<td>one of &quot;Wang&quot;, &quot;Resnik&quot;, &quot;Rel&quot;, &quot;Jiang&quot;, and &quot;Lin&quot;.</td>
</tr>
</tbody>
</table>

**Details**

provide two DO term vectors, this function will calculate their similarities.

**Value**

score matrix

Author(s)

Guangchuang Yu [http://ygc.name](http://ygc.name)
Description

dotplot method

Usage

dotplot(object, ...)

## S4 method for signature 'enrichResult'
dotplot(object, x = "geneRatio",
        colorBy = "p.adjust", showCategory = 10, font.size = 12, title = "")

Arguments

object an instance of enrichResult
...
additional parameter
x variable for x axis
colorBy one of 'pvalue', 'p.adjust' and 'qvalue'
showCategory number of category
font.size font size
title plot title

Value

plot

Author(s)

Guangchuang Yu
Guangchuang Yu
Description

Intral method for enrichment analysis

Usage

enrich.internal(gene, organism, pvalueCutoff, pAdjustMethod = "BH", ont, universe, minGSSize = 5, qvalueCutoff = 0.2, readable = FALSE, ...)

Arguments

gene a vector of entrez gene id.
organism supported organism.
pvalueCutoff Cutoff value of pvalue.
pAdjustMethod one of "holm", "hochberg", "hommel", "bonferroni", "BH", "BY", "fdr", "none"
ont Ontology
universe background genes
minGSSize minimal size of genes annotated by Ontology term for testing.
qvalueCutoff cutoff of qvalue
readable whether mapping gene ID to gene Name
... additional parameter

Details

using the hypergeometric model

Value

A enrichResult instance.

Author(s)

Guangchuang Yu http://ygc.name
enrichDO  DO Enrichment Analysis

Description
Given a vector of genes, this function will return the enrichment DO categories with FDR control.

Usage
enrichDO(gene, ont = "DO", pvalueCutoff = 0.05, pAdjustMethod = "BH", universe, minGSSize = 5, qvalueCutoff = 0.2, readable = FALSE)

Arguments
- gene: a vector of entrez gene id.
- ont: one of DO or DOLite.
- pvalueCutoff: Cutoff value of pvalue.
- pAdjustMethod: one of "holm", "hochberg", "hommel", "bonferroni", "BH", "BY", "fdr", "none"
- universe: background genes
- minGSSize: minimal size of genes annotated by Ontology term for testing.
- qvalueCutoff: qvalue Cutoff
- readable: whether mapping gene ID to gene Name

Value
A enrichResult instance.

Author(s)
Guangchuang Yu http://ygc.name

See Also
enrichResult-class

Examples
data(geneList)
gene = names(geneList)[geneList > 1]
yy = enrichDO(gene, pvalueCutoff=0.05)
summary(yy)
enrichMap

Description
enrichment map

Usage
enrichMap(x, n = 50, fixed = TRUE, vertex.label.font = 1, ...)

Arguments
- x: gseaResult or enrichResult object
- n: maximum number of category to shown
- fixed: if set to FALSE, will invoke tkplot
- vertex.label.font: font size of vertex label
- ...: additional parameter

Details
enrichment map

Value
figure

Author(s)
G Yu

enrichNCG

Description
Enrichment analysis based on the Network of Cancer Genes database (http://ncg.kcl.ac.uk/)

Usage
enrichNCG(gene, pvalueCutoff = 0.05, pAdjustMethod = "BH", universe, minGSSize = 5, qvalueCutoff = 0.2, readable = FALSE)
Arguments

gene a vector of entrez gene id
pvalueCutoff pvalue cutoff
pAdjustMethod one of "holm", "hochberg", "hommel", "bonferroni", "BH", "BY", "fdr", "none"
universe background genes
minGSSize minimal size of genes annotated by NCG category for testing
qvalueCutoff qvalue cutoff
readable whether mapping gene ID to gene Name

Details

given a vector of genes, this function will return the enrichment NCG categories with FDR control

Value

A enrichResult instance

Author(s)

Guangchuang Yu

enrichResult-class

Class "enrichResult" This class represents the result of enrichment analysis.

Description

Class "enrichResult" This class represents the result of enrichment analysis.

Slots

result enrichment analysis
pvalueCutoff pvalueCutoff
pAdjustMethod pvalue adjust method
qvalueCutoff qvalue cutoff
organism only "human" supported
ontology biological ontology
gene Gene IDs
universe background gene
geneInCategory gene and category association
geneSets gene sets
readable logical flag of gene ID in symbol or not.
**EXTID2NAME**

**Description**

mapping gene ID to gene Symbol

**Usage**

`EXTID2NAME(geneID, organism)`

**Arguments**

- **geneID**: entrez gene ID
- **organism**: one of "human", "mouse" and "yeast"

**Value**

gene symbol

**Author(s)**

Guangchuang Yu [http://ygc.name](http://ygc.name)

---

**EXTID2TERMID**

**Description**

Mapping External ID to Ontology Term ID

**Usage**

`EXTID2TERMID(gene, organism, ...)`

**Arguments**

- **gene**: gene ID vector
- **organism**: organism
- **...**: additional parameter
Description

fortify

Usage

## S3 method for class 'enrichResult'
fortify(model, data, showCategory = 5, order = FALSE, 
drop = FALSE, ...)

Arguments

- model: enrichResult object
- data: not use here
- showCategory: Category numbers to show
- order: logical
- drop: logical
- ...: additional parameter

Value

figure
gene2DO

Author(s)
G Yu

gene2DO  
convert Gene ID to DO Terms

Description
provide gene ID, this function will convert to the corresponding DO Terms

Usage
gene2DO(gene)

Arguments
gene  
etrez gene ID

Value
DO Terms

Author(s)
Guangchuang Yu http://ygc.name

geneSim  
geneSim

Description
measuring similarities between two gene vectors.

Usage
geneSim(geneID1, geneID2 = NULL, measure = "Wang", combine = "BMA")

Arguments
geneID1  
etrez gene vector
geneID2  
etrez gene vector
measure  
one of "Wang", "Resnik", "Rel", "Jiang", and "Lin".
combine  
One of "max", "average", "rcmax", "BMA" methods, for combining semantic similarity scores of multiple DO terms associated with gene/protein.
Details

provide two entrez gene vectors, this function will calculate their similarity.

Value

score matrix

Author(s)

Guangchuang Yu http://ygc.name

Description

get all entrezgene ID of a specific organism

Usage

getALLEG(organism)

Arguments

organism species

Value

etrez gene ID vector

Author(s)

Yu Guangchuang
Description
preparing geneSets for gene set enrichment analysis

Usage
getGeneSet(setType, organism, ...)

Arguments
setType type of gene sets
organism organism
... additional parameter

gsea

description
gene set enrichment analysis

Usage
gsea(geneList, geneSets, setType, organism, exponent, nPerm, minGSSize,
pvalueCutoff, pAdjustMethod, verbose, seed = FALSE, ...)

Arguments
geneList order ranked geneList
geneSets gene sets
setType Type of geneSet
organism organism
exponent weight of each step
nPerm permutation numbers
minGSSize minimal size of each geneSet for analyzing
pvalueCutoff p value Cutoff
pAdjustMethod p value adjustment method
verbose print message or not
seed set seed inside the function to make result reproducible. FALSE by default.
... additional parameter
gseAnalyzer

**Value**

gseaResult object

**Author(s)**

Yu Guangchuang

---

**gseAnalyzer** | **Gene Set Enrichment Analysis**

**Description**

perform gsea analysis

**Usage**

gseAnalyzer(geneList, setType, organism = "human", exponent = 1, nPerm = 1000, minGSSize = 10, pvalueCutoff = 0.05, pAdjustMethod = "BH", verbose = TRUE, ...)

**Arguments**

- **geneList**: order ranked geneList
- **setType**: Type of geneSet
- **organism**: organism
- **exponent**: weight of each step
- **nPerm**: permutation numbers
- **minGSSize**: minimal size of each geneSet for analyzing
- **pvalueCutoff**: pvalue Cutoff
- **pAdjustMethod**: p value adjustment method
- **verbose**: print message or not
- **...**: additional parameters

**Value**

gseaResult object

**Author(s)**

Yu Guangchuang
gseaplot

visualize analyzing result of GSEA

Description
plotting function for gseaResult

Usage
gseaplot(gseaResult, geneSetID, by = "all")

Arguments

  gseaResult          gseaResult object
  geneSetID           geneSet ID
  by                  one of "runningScore" or "position"

Value

  ggplot2 object

Author(s)
Yu Guangchuang

---

gseaResult-class

Class "gseaResult" This class represents the result of GSEA analysis

Description

Class "gseaResult" This class represents the result of GSEA analysis

Slots

  result  GSEA analysis
  setType  setType
  geneSets  geneSets
  geneList  order rank geneList
  permScores  permutation scores
  params  parameters

Author(s)

  Guangchuang Yu http://ygc.name
See Also
gseaplot

list2graph

**Description**

convert a list of gene IDs to igraph object.

**Usage**

list2graph(inputList)

**Arguments**

inputList a list of gene IDs

**Value**

a igraph object.

**Author(s)**

Guangchuang Yu [http://yg.c.name](http://yg.c.name)

mclusterSim

**Description**

Pairwise semantic similarity for a list of gene clusters

**Usage**

mclusterSim(clusters, measure = "Wang", combine = "BMA")

**Arguments**

clusters A list of gene clusters
measure one of "Wang", "Resnik", "Rel", "Jiang", and "Lin".
combine One of "max", "average", "rcmax", "BMA" methods, for combining semantic similarity scores of multiple DO terms associated with gene/protein.
**Description**
plot network

**Usage**
```r
netplot(g, vertex.label.font = 2, vertex.label.color = "#666666",
        vertex.label.cex = 1.5, layout = layout.fruchterman.reingold,
        foldChange = NULL, fixed = TRUE, col.bin = 10, legend.x = 1,
        legend.y = 1)
```

**Arguments**
- `g` igraph object
- `vertex.label.font` font size
- `vertex.label.color` font text color
- `vertex.label.cex` cex of vertex label
- `layout` layout
- `foldChange` fold change
- `fixed` logical
- `col.bin` number of legend color bin
- `legend.x` x-axis position of legend
- `legend.y` y-axis position of legend
Details

plot network of igraph object

Value

plot

Author(s)

Yu Guangchuang

Description

plot method generics
plot method for gseaResult

Usage

```r
## S4 method for signature 'enrichResult,ANY'
plot(x, type = "bar", ...)

## S4 method for signature 'gseaResult,ANY'
plot(x, type = "gseaplot", ...)
```

Arguments

- `x` A enrichResult instance
- `type` one of bar, cnet or enrichMap
- `...` Additional argument list

Value

plot

plot

Author(s)

Guangchuang Yu [http://ygc.name](http://ygc.name)

Yu Guangchuang
rebuildAnnoData  

rebuilding annotation data

Description
rebuilding entrez and DO mapping datasets

Usage
rebuildAnnoData(file)

Arguments
file do_rif.human.txt

Author(s)
Guangchuang Yu http://ygc.name

scaleNodeColor  

scaleNodeColor

Description
scale color nodes

Usage
scaleNodeColor(g, foldChange, node.idx = NULL, DE.foldChange)

Arguments
g igraph object
foldChange fold Change
node.idx index of node to color
DE.foldChange logical

Details
color nodes based on fold change of expression

Value
igraph object

Author(s)
Yu Guangchuang
Description
mapping geneID to gene Symbol

Usage
setReadable(x)

Arguments
x enrichResult Object

Value
enrichResult Object

Author(s)
Yu Guangchuang

Description
setting basic attributes of a graph

Usage
setting.graph.attributes(g, node.size = 8, node.color = "#B3B3B3", edge.width = 2, edge.color = "#8DA0CB")

Arguments
g igraph object
node.size size of node
node.color color of node
dgege.width edge width
dgege.color color of edge
show

Details

setting size and color of node and edge

Value

igraph object

Author(s)

Yu Guangchuang

---

show method

Description

show method for enrichResult instance
show method for gseaResult instance

Usage

show(object)

Arguments

object A enrichResult instance.

Value

message

Author(s)

Guangchuang Yu http://ygc.name
Guangchuang Yu http://ygc.name
Description

plotting similarity matrix

Usage

```r
simplot(sim, xlab = ",", ylab = ",", color.low = "white",
color.high = "red", labs = TRUE, digits = 2, labs.size = 3,
font.size = 14, readable = FALSE)
```

Arguments

- `sim`: similarity matrix
- `xlab`: xlab
- `ylab`: ylab
- `color.low`: color of low value
- `color.high`: color of high value
- `labs`: logical, add text label or not
- `digits`: round digit numbers
- `labs.size`: lable size
- `font.size`: font size
- `readable`: TRUE or FALSE

Value

- ggplot object

Author(s)

Yu Guangchuang
**summary**

**summary method**

*Description*

summary method for enrichResult instance
summary method for gseaResult instance

*Usage*

summary(object, ...)

summary(object, ...)

*Arguments*

object A enrichResult instance.
...
additional parameter

*Value*

A data frame
A data frame

*Author(s)*

Guangchuang Yu [http://ygc.name](http://ygc.name)
Guangchuang Yu [http://ygc.name](http://ygc.name)

**TERM2NAME**

*TERM2NAME*

*Description*

Mapping Ontology Term ID to Name Symbol or Description

*Usage*

TERM2NAME(term, organism, ...)

*Arguments*

term term ID vector
organism organism
...
additional parameter
Description
Mapping Ontology Term ID to External ID

Usage
TERMID2EXTID(term, organism, ...)

Arguments
<table>
<thead>
<tr>
<th>term</th>
<th>term ID vector</th>
</tr>
</thead>
<tbody>
<tr>
<td>organism</td>
<td>organism</td>
</tr>
<tr>
<td>...</td>
<td>additional parameter</td>
</tr>
</tbody>
</table>

Description
ggplot theme of DOSE

Usage
theme_dose(font.size = 14)

Arguments
| font.size | font size |
Index

*Topic classes
  enrichResult-class, 12
  gseaResult-class, 19
*Topic datasets
  DataSet, 7
*Topic manip
  enrich.internal, 9
  enrichDO, 10
  gseAnalyzer, 18
*Topic package
  DOSE-package, 3
  ALLEXTID, 3
  barplot.enrichResult, 4
  clusterSim, 4
  cnetplot, 5
  cnetplot.enrichResult-method (cnetplot), 5
  cnetplot.internal, 6
  computeIC, 6
  DataSet, 7
  D02ALLEG(DataSet), 7
  D02EG(DataSet), 7
  DOLite2EG(DataSet), 7
  DOLiteTerm(DataSet), 7
  DOSE (DOSE-package), 3
  DOSE-package, 3
  DOSEEEnv(DataSet), 7
  doSim, 7
  dotplot, 8
  dotplot.enrichResult, ANY-method (dotplot), 8
  dotplot.enrichResult-method (dotplot), 8
  EG2ALLDO(DataSet), 7
  EG2DO(DataSet), 7
  EG2DOLite(DataSet), 7
  enrich.internal, 9
  enrichDO, 10, 13
  enrichMap, 11
  enrichNCG, 11
  enrichResult, 3
  enrichResult-class, 12
  EXTID2NAME, 13
  EXTID2TERMID, 13
  fortify.enrichResult, 14
  fortify.gseaResult, 14
  gene2DO, 15
  geneList(DataSet), 7
  geneSim, 15
  getALLEG, 16
  getGeneSet, 17
  gsea, 17
  gseahResult-class (gseaResult-class), 19
  gseAnalyzer, 18
  gseaplot, 19, 20
  gseaResult-class, 19
  IC (DataSet), 7
  Info Contents_human_DO (DataSet), 7
  list2graph, 20
  mclusterSim, 20
  NCG_DOSE_Env (DataSet), 7
  netplot, 21
  plot, 22
  plot.enrichResult, ANY-method (plot), 22
  plot.enrichResult-method (enrichResult-class), 12
  plot, gseaResult, ANY-method (plot), 22
  plot, gseaResult-method (gseaResult-class), 19
  rebuildAnnoData, 23
scaleNodeColor, 23
setReadable, 24
setting.graph.attributes, 24
show, 25
show, enrichResult-method
   (enrichResult-class), 12
show, gseaResult-method
   (gseaResult-class), 19
simplot, 26
summary, 27
summary, enrichResult-method
   (enrichResult-class), 12
summary, gseaResult-method
   (gseaResult-class), 19

TERM2NAME, 27
TERMID2EXTID, 28
theme_dose, 28