Package 'GOSemSim'

March 26, 2013

Type Package

Title GO-terms Semantic Similarity Measures
Version 1.16.1
Author Guangchuang Yu
Maintainer Guangchuang Yu < guangchuangyu@gmail.com>
Description Implemented five methods proposed by Resnik, Schlicker, Jiang, Lin and Wang respectively for estimating GO semantic similarities. Support many species, including Anopheles, Arabidopsis, Bovine, Canine, Chicken, Chimp, Coelicolor, E coli strain K12 and Sakai, Fly, Human, Malaria, Mouse, Pig,Rhesus, Rat, Worm, Xenopus, Yeast, and Zebrafish.
Depends R (>= 2.10), Rcpp
LinkingTo Rcpp
Imports methods, AnnotationDbi, GO.db, org.Hs.eg.db
Suggests DOSE, clusterProfiler
biocViews GO, Clustering, Pathways, NetworkAnalysis
Collate 'clusterSim.R' 'combineMethods.R' 'computeIC.R' 'ICMethods.R' 'gene2GO.R' 'geneSim.R' 'goSim.R' 'termSim.R package.R' 'mgeneSim.R' 'mgoSim.R' 'mclusterSim.R'
License GPL-2
$\textbf{URL} \ \text{http://bioinformatics.oxfordjournals.org/content/} \\ 26/7/976.full$
R topics documented:

GOS	emSim-package	9		G	en	e	Oı	nte	ole	gy	y-l	bas	sec	l S	Sei	na	ıti	c S	Sin	nil	ar	ity	, N	1e	as	ur	es					
Index																																12
	wangMethod	•						•					٠	٠			•			•	٠	•	•							 	•	11
	termSim																															
	mgoSim																															
	mgeneSim																															
	mclusterSim .																															

Description

Implementation of semantic similarity measures to estimate the functional similarities among Gene Ontology terms and gene products

Details

Quantitative measure of functional similarities among gene products is important for post-genomics study. and widely used in gene function prediction, cluster analysis and pathway modeling. This package is designed to estimate the GO terms' and genes' semantic similarities. Implemented five methods proposed by Resnik, Schlicker, Jiang, Lin and Wang respectively. Support many species, including Anopheles, Arabidopsis, Bovine, Canine, Chicken, Chimp, E coli strain K12 and strain Sakai, Fly, Human, Malaria, Mouse, Pig, Rhesus, Rat, Worm, Xenopus, Yeast, Zebrafish.

Package: GOSemSim Type: Package Version: 1.6.0 Date: 09-11-2012

biocViews: GO, Clustering, Pathways, Anopheles_gambiae, Arabidopsis_thaliana, Bos_taurus, Caenorhabditis_elegans

Depends:

Imports: methods, AnnotationDbi, GO.db, org.Hs.eg.db, org.Ag.eg.db, org.At.tair.db, org.Bt.eg.db, org.Ce.eg.db, org

Suggests: clusterProfiler License: GPL Version 2

Author(s)

Guangchuang Yu

Maintainer: Guangchuang Yu <guangchuangyu@gmail.com>

References

Yu et al. (2010) GOSemSim: an R package for measuring semantic similarity among GO terms and gene products *Bioinformatics* (Oxford, England), 26:7 976–978, April 2010. ISSN 1367-4803 http://bioinformatics.oxfordjournals.org/cgi/content/abstract/26/7/976 PMID: 20179076

See Also

goSim mgoSim geneSim mgeneSim clusterSim mclusterSim

clusterSim 3

1	ı , a.
C	lusterSim

Semantic Similarity Between Two Gene Clusters

Description

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

Usage

```
clusterSim(cluster1, cluster2, ont = "MF", organism = "human", measure = "Wang", drop = "IEA", combine = "BMA")
```

Arguments

cluster1 A set of gene IDs.
cluster2 Another set of gene IDs.

ont One of "MF", "BP", and "CC" subontologies.

measure One of "Resnik", "Lin", "Rel", "Jiang" and "Wang" methods.

organism One of "anopheles", "arabidopsis", "bovine", "canine", "chicken", "chimp", "coeli-

color", "ecolik12", "ecsakai", "fly", "human", "malaria", "mouse", "pig", "rat",

"rhesus", "worm", "xenopus", "yeast" and "zebrafish".

drop A set of evidence codes based on which certain annotations are dropped. Use

NULL to keep all GO annotations.

combine One of "max", "average", "rcmax", "BMA" methods, for combining semantic

similarity scores of multiple GO terms associated with protein or multiple pro-

teins assiciated with protein cluster.

Value

similarity

References

Yu et al. (2010) GOSemSim: an R package for measuring semantic similarity among GO terms and gene products *Bioinformatics* (Oxford, England), 26:7 976–978, April 2010. ISSN 1367-4803 http://bioinformatics.oxfordjournals.org/cgi/content/abstract/26/7/976 PMID: 20179076

See Also

goSim mgoSim geneSim mgeneSim mclusterSim

Examples

4 geneSim

	1 .	α	
com	hin	P.50	cores

combining similarity matrix to similarity score

Description

Functions for combining similarity matrix to similarity score

Usage

```
combineScores(SimScores, combine)
```

Arguments

SimScores similarity matrix combine combine method

Value

similarity value

Author(s)

Guangchuang Yu http://ygc.name

geneSim

Semantic Similarity Between two Genes

Description

Given two genes, this function will calculate the semantic similarity between them, and return their semantic similarity and the corresponding GO terms

Usage

```
geneSim(gene1, gene2, ont = "MF", organism = "human", measure = "Wang", drop = "IEA", combine = "BMA")
```

Arguments

combine

gene1	Entrez gene id.
gene2	Another entrez gene id.
ont	One of "MF", "BP", and "CC" subontologies.
organism	One of "anopheles", "arabidopsis", "bovine", "canine", "chicken", "chimp", "coelicolor", "ecolik12", "ecsakai", "fly", "human", "malaria", "mouse", "pig", "rat", "rhesus", "worm", "xenopus", "yeast" and "zebrafish".
measure	One of "Resnik", "Lin", "Rel", "Jiang" and "Wang" methods.
drop	A set of evidence codes based on which certain annotations are dropped. Use NULL to keep all GO annotations.

One of "max", "average", "rcmax", "BMA" methods, for combining semantic

similarity scores of multiple GO terms associated with protein or multiple pro-

teins assiciated with protein cluster.

goSim 5

Value

list of similarity value and corresponding GO.

References

Yu et al. (2010) GOSemSim: an R package for measuring semantic similarity among GO terms and gene products *Bioinformatics* (Oxford, England), 26:7 976–978, April 2010. ISSN 1367-4803 http://bioinformatics.oxfordjournals.org/cgi/content/abstract/26/7/976 PMID: 20179076

See Also

 $goSim\ mgoSim\ mgeneSim\ clusterSim\ mclusterSim$

Examples

```
geneSim("241", "251", ont="MF", organism="human", measure="Wang")
```

goSim

Semantic Similarity Between Two GO Terms

Description

Given two GO IDs, this function calculates their semantic similarity.

Usage

```
goSim(GOID1, GOID2, ont = "MF", organism = "human", measure = "Wang")
```

Arguments

GOID1 GO ID 1. GOID2 GO ID 2.

ont One of "MF", "BP", and "CC" subontologies.

organism One of "anopheles", "arabidopsis", "bovine", "canine", "chicken", "chimp", "coeli-

color", "ecolik12", "ecsakai", "fly", "human", "malaria", "mouse", "pig", "rat", "rhesus",

"worm", "xenopus", "yeast" and "zebrafish".

measure One of "Resnik", "Lin", "Rel", "Jiang" and "Wang" methods.

Value

similarity

References

Yu et al. (2010) GOSemSim: an R package for measuring semantic similarity among GO terms and gene products *Bioinformatics* (Oxford, England), 26:7 976–978, April 2010. ISSN 1367-4803 http://bioinformatics.oxfordjournals.org/cgi/content/abstract/26/7/976 PMID: 20179076

6 infoContentMethod

See Also

mgoSim geneSim mgeneSim clusterSim mclusterSim

Examples

```
goSim("GO:0004022", "GO:0005515", ont="MF", measure="Wang")
```

IC

Information content of GO terms

Description

These datasets are the information contents of GOterms.

References

Yu et al. (2010) GOSemSim: an R package for measuring semantic similarity among GO terms and gene products *Bioinformatics* (Oxford, England), 26:7 976–978, April 2010. ISSN 1367-4803 http://bioinformatics.oxfordjournals.org/cgi/content/abstract/26/7/976 PMID: 20179076

infoContentMethod

information content based methods

Description

Information Content Based Methods for semantic similarity measuring

Usage

```
infoContentMethod(ID1, ID2, ont = "DO", method, organism = "human")
```

Arguments

ID1 Ontology TermID2 Ontology Termont Ontology

method one of "Resnik", "Jiang", "Lin" and "Rel".

organism one of supported species

Details

implemented for methods proposed by Resnik, Jiang, Lin and Schlicker.

Value

semantic similarity score

Author(s)

Guangchuang Yu http://ygc.name

loadICdata 7

loadICdata Load IC data

Description

Load Information Content data to DOSEEnv environment

Usage

loadICdata(organism, ont)

Arguments

organism "human" ont "DO"

Value

NULL

Author(s)

Guangchuang Yu http://ygc.name

mclusterSim

Pairwise Semantic Similarities for a List of Gene Clusters

Description

Given a list of gene clusters, this function calculates pairwise semantic similarities.

Usage

```
mclusterSim(clusters, ont = "MF", organism = "human", measure = "Wang", drop = "IEA", combine = "BMA")
```

Arguments

clusters A list of gene clusters.

ont One of "MF", "BP", and "CC" subontologies.

 $\label{eq:measure} \qquad \qquad \text{One of "Resnik", "Lin", "Rel", "Jiang" and "Wang" methods.}$

organism One of "anopheles", "arabidopsis", "bovine", "canine", "chicken", "chimp", "coeli-

color", "ecolik12", "ecsakai", "fly", "human", "malaria", "mouse", "pig", "rat",

"rhesus", "worm", "xenopus", "yeast" and "zebrafish".

drop A set of evidence codes based on which certain annotations are dropped. Use

NULL to keep all GO annotations.

combine One of "max", "average", "rcmax", "BMA" methods, for combining semantic

similarity scores of multiple GO terms associated with protein or multiple pro-

teins assiciated with protein cluster.

8 mgeneSim

Value

similarity matrix

References

Yu et al. (2010) GOSemSim: an R package for measuring semantic similarity among GO terms and gene products *Bioinformatics* (Oxford, England), 26:7 976–978, April 2010. ISSN 1367-4803 http://bioinformatics.oxfordjournals.org/cgi/content/abstract/26/7/976 PMID: 20179076

See Also

goSim mgoSim geneSim mgeneSim clusterSim

Examples

mgeneSim

Pairwise Semantic Similarity for a List of Genes

Description

Given a list of genes, this function calculates pairwise semantic similarities.

Usage

```
mgeneSim(genes, ont = "MF", organism = "human",
measure = "Wang", drop = "IEA", combine = "BMA",
verbose = TRUE)
```

Arguments

A list of entrez gene IDs. genes One of "MF", "BP", and "CC" subontologies. ont One of "Resnik", "Lin", "Rel", "Jiang" and "Wang" methods. measure One of "anopheles", "arabidopsis", "bovine", "canine", "chicken", "chimp", "coeliorganism color", "ecolik12", "ecsakai", "fly", "human", "malaria", "mouse", "pig", "rat", "rhesus", "worm", "xenopus", "yeast" and "zebrafish". drop A set of evidence codes based on which certain annotations are dropped. Use NULL to keep all GO annotations. One of "max", "average", "rcmax", "BMA" methods, for combining semantic combine similarity scores of multiple GO terms associated with protein or multiple proteins assiciated with protein cluster. verbose show progress bar or not.

mgoSim 9

Value

similarity matrix

References

Yu et al. (2010) GOSemSim: an R package for measuring semantic similarity among GO terms and gene products *Bioinformatics* (Oxford, England), 26:7 976–978, April 2010. ISSN 1367-4803 http://bioinformatics.oxfordjournals.org/cgi/content/abstract/26/7/976 PMID: 20179076

See Also

 $goSim\ mgoSim\ geneSim\ clusterSim\ mclusterSim$

Examples

```
mgeneSim(c("835", "5261", "241"), ont="MF", organism="human", measure="Wang")
```

mgoSim

Semantic Similarity Between two GO terms lists

Description

Given two GO term sets, this function will calculate the semantic similarity between them, and return their semantic similarity

Usage

```
mgoSim(GO1, GO2, ont = "MF", organism = "human", measure = "Wang", combine = "BMA")
```

teins assiciated with protein cluster.

Arguments

GO1	A set of go terms.
GOI	Tibet of go terms.
GO2	Another set of go terms.
ont	One of "MF", "BP", and "CC" subontologies.
organism	One of "anopheles", "arabidopsis", "bovine", "canine", "chicken", "chimp", "coelicolor", "ecolik12", "ecsakai", "fly", "human", "malaria", "mouse", "pig", "rat", "rhesus", "worm", "xenopus", "yeast" and "zebrafish".
measure	One of "Resnik", "Lin", "Rel", "Jiang" and "Wang" methods.
combine	One of "max", "average", "rcmax", "BMA" methods, for combining semantic

similarity scores of multiple GO terms associated with protein or multiple pro-

Value

similarity

10 termSim

References

Yu et al. (2010) GOSemSim: an R package for measuring semantic similarity among GO terms and gene products *Bioinformatics* (Oxford, England), 26:7 976–978, April 2010. ISSN 1367-4803 http://bioinformatics.oxfordjournals.org/cgi/content/abstract/26/7/976 PMID: 20179076

See Also

 ${\tt goSim\ geneSim\ mgeneSim\ clusterSim\ mclusterSim}$

Examples

```
\begin{array}{l} go1 <- c("GO:0004022", "GO:0004024", "GO:0004023") \\ go2 <- c("GO:0009055", "GO:0020037") \\ mgoSim("GO:0003824", go2, measure="Wang") \\ mgoSim(go1, go2, ont="MF", organism="human", measure="Wang") \\ \end{array}
```

 $\operatorname{termSim}$

termSim

Description

measuring similarities between two term vectors.

Usage

```
termSim(t1, t2, method = "Wang", organism = "human", ont)
```

Arguments

```
t1 term vector
t2 term vector
method one of "Wang", "Resnik", "Rel", "Jiang", and "Lin".
organism only "human" supported
ont ontology
```

Details

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

Value

score matrix

Author(s)

Guangchuang Yu http://ygc.name

wangMethod 11

$\mathbf{wangMethod}$	wangMethod
-----------------------	------------

Description

Method Wang for semantic similarity measuring

Usage

```
wangMethod(ID1, ID2, ont)
```

Arguments

ID1	Ontology Term
ID2	Ontology Term
ont	Ontology

Value

semantic similarity score

Author(s)

Guangchuang Yu http://ygc.name

Index

*Topic datasets	Info_Contents_ecolik12_MF (IC), 6
IC, 6	Info_Contents_ecsakai_BP (IC), 6
*Topic manip	Info_Contents_ecsakai_CC (IC), 6
clusterSim, 3	Info_Contents_ecsakai_MF (IC), 6
geneSim, 4	Info_Contents_fly_BP (IC), 6
goSim, 5	Info_Contents_fly_CC (IC), 6
mclusterSim, 7	Info_Contents_fly_MF (IC), 6
mgeneSim, 8	Info_Contents_human_BP (IC), 6
mgoSim, 9	Info_Contents_human_CC (IC), 6
*Topic package	Info_Contents_human_MF (IC), 6
GOSemSim-package, 2	Info_Contents_malaria_BP (IC), 6
1 0 /	Info Contents malaria CC (IC), 6
clusterSim, 2, 3, 5, 6, 8–10	Info Contents malaria MF (IC), 6
combineScores, 4	Info Contents mouse BP (IC), 6
	Info_Contents_mouse_CC (IC), 6
geneSim, 2, 3, 4, 6, 8–10	Info Contents mouse MF (IC), 6
GOSemSim (GOSemSim-package), 2	Info_Contents_pig_BP (IC), 6
GOSemSim-package, 2	Info_Contents_pig_CC (IC), 6
goSim, 2, 3, 5, 5, 8–10	Info_Contents_pig_MF (IC), 6
	Info Contents rat BP (IC), 6
IC, 6	Info Contents rat CC (IC), 6
Info_Contents_anopheles_BP (IC), 6	Info Contents rat MF (IC), 6
Info_Contents_anopheles_CC (IC), 6	Info Contents rhesus BP (IC), 6
Info_Contents_anopheles_MF (IC), 6	Info_Contents_rhesus_CC (IC), 6
Info_Contents_arabidopsis_BP (IC), 6	Info_Contents_rhesus_MF (IC), 6
Info_Contents_arabidopsis_CC (IC), 6	Info Contents worm BP (IC), 6
Info_Contents_arabidopsis_MF (IC), 6	Info Contents worm CC (IC), 6
Info_Contents_bovine_BP (IC), 6	Info Contents worm MF (IC), 6
Info_Contents_bovine_CC (IC), 6	Info Contents xenopus BP (IC), 6
Info_Contents_bovine_MF (IC), 6	Info Contents xenopus CC (IC), 6
Info_Contents_canine_BP (IC), 6	Info Contents xenopus MF (IC), 6
Info_Contents_canine_CC (IC), 6	Info_Contents_yeast_BP (IC), 6
Info_Contents_canine_MF (IC), 6	Info Contents yeast CC (IC), 6
Info_Contents_chicken_BP (IC), 6	Info_Contents_yeast_MF (IC), 6
Info_Contents_chicken_CC (IC), 6	Info Contents zebrafish BP (IC), 6
Info_Contents_chicken_MF (IC), 6	Info Contents zebrafish CC (IC), 6
Info_Contents_chimp_BP (IC), 6	Info Contents zebrafish MF (IC), 6
Info_Contents_chimp_CC (IC), 6	infoContentMethod, 6
Info_Contents_chimp_MF (IC), 6	
Info_Contents_coelicolor_BP (IC), 6	loadICdata, 7
Info_Contents_coelicolor_CC (IC), 6	
Info_Contents_coelicolor_MF (IC), 6	mclusterSim, 2, 3, 5, 6, 7, 9, 10
Info_Contents_ecolik12_BP (IC), 6	mgeneSim, 2, 3, 5, 6, 8, 8, 10
Info_Contents_ecolik12_CC (IC), 6	mgoSim, 2, 3, 5, 6, 8, 9, 9

INDEX 13

 $\mathrm{termSim}, \frac{10}{}$

 ${\rm wangMethod}, 11$