# Package 'GOSemSim'

October 12, 2016

Type Package
Title GO-terms Semantic Similarity Measures
Version 1.30.3
<b>Author</b> Guangchuang Yu <guangchuangyu@gmail.com> with contributions from Alexey Stukalov and Chuanle Xiao.</guangchuangyu@gmail.com>
Maintainer Guangchuang Yu <guangchuangyu@gmail.com></guangchuangyu@gmail.com>
<b>Description</b> 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, Gondii, Human, Malaria, Mouse, Pig, Rhesus, Rat, Worm, Xenopus, Yeast, and Zebrafish.
<b>Depends</b> R (>= $3.2.0$ )
LinkingTo Rcpp
Imports Rcpp, AnnotationDbi, GO.db, utils
Suggests DOSE, clusterProfiler, org.Hs.eg.db, knitr, BiocStyle, BiocInstaller
VignetteBuilder knitr
License Artistic-2.0
<pre>URL http://guangchuangyu.github.io/GOSemSim</pre>
<pre>BugReports https://github.com/GuangchuangYu/GOSemSim/issues</pre>
biocViews Annotation, GO, Clustering, Pathways, Network, Software
RoxygenNote 5.0.1
NeedsCompilation yes
R topics documented:
GOSemSim-package

	geneSim																										
	getDb																										6
	getSupported_Org .																										6
	goSim																										7
	IC																										8
	infoContentMethod																										8
	loadGOMap																										9
	loadICdata																										9
	mclusterSim																										10
	mgeneSim																										11
	mgoSim																										12
	termSim																										13
	wangMethod																								 		14
Index																											15
GOSem	nSim-package	Gen	e O	nte	olo	уду	-ba	ıse	d l	Sei	ma	ıtio	c S	im	iile	ari	ity	M	ea	su	ıre	?S					

## **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: 1.24.0 Date: 09-11-2012

biocViews: GO, Clustering, Pathways, Anopheles\_gambiae, Arabidopsis\_thaliana, Bos\_taurus, Caenorhabditis\_elegans, Ca

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.Cf.e

Suggests: clusterProfiler License: Artistic-2.0

#### Author(s)

Guangchuang Yu

Maintainer: Guangchuang Yu <guangchuangyu@gmail.com>

clusterSim 3

#### 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

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. One of "MF", "BP", and "CC" subontologies. ont One of "anopheles", "arabidopsis", "bovine", "canine", "chicken", "chimp", "coeliorganism color", "ecolik12", "ecsakai", "fly", "gondii", "human", "malaria", "mouse", "pig", "rat", "rhesus", "worm", "xenopus", "yeast" and "zebrafish". One of "Resnik", "Lin", "Rel", "Jiang" and "Wang" methods. measure 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.

## 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

4 combineScores

#### See Also

goSim mgoSim geneSim mgeneSim mclusterSim

## **Examples**

```
## cluster1 <- c("835", "5261","241", "994")
## cluster2 <- c("307", "308", "317", "321", "506", "540", "378", "388", "396")
## clusterSim(cluster1, cluster2, ont="MF", organism="human", measure="Wang")</pre>
```

combineScores

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 5

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**

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", "gondii", "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.
combine	One of "max", "average", "rcmax", "BMA" methods, for combining semantic similarity scores of multiple GO terms associated with protein or multiple proteins assiciated with protein cluster.

#### 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")
```

6 getSupported\_Org

getDb

getDb

## Description

mapping organism name to annotationDb package name

# Usage

getDb(organism)

## Arguments

 ${\tt organism}$ 

one of supported organism

#### Value

annotationDb name

## Author(s)

Yu Guangchuang

getSupported\_Org

getSupported\_Org

# Description

get supported organisms

# Usage

getSupported\_Org()

## Value

supported organisms

## Author(s)

Yu Guangchuang

goSim 7

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

#### See Also

mgoSim geneSim mgeneSim clusterSim mclusterSim

## **Examples**

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

8 infoContentMethod

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 = "D0", method, organism = "human")
```

## Arguments

ID1	Ontology Term
ID2	Ontology Term
	0.4.1

ont 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

loadGOMap 9

loadGOMap

load GOMap

## Description

loading GOMap to GOSemSimEnv

## Usage

loadGOMap(organism)

## Arguments

organism

one of supported organisms

#### Value

envir

#### Author(s)

Yu Guangchuang

loadICdata

Load IC data

# Description

Load Information Content data to DOSEEnv environment

# Usage

```
loadICdata(organism, ont)
```

# Arguments

organism "human" ont "DO"

## Author(s)

Guangchuang Yu http://ygc.name

10 mclusterSim

mclusterSim	Pairwise Semantic Similarities for a List of Gene Clusters
merascer sim	Tuil Wise Semantic Stitute West for a Bist of Gene Constens

## **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.
organism	One of "anopheles", "arabidopsis", "bovine", "canine", "chicken", "chimp", "coelicolor", "ecolik12", "ecsakai", "fly", "gondii", "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.
combine	One of "max", "average", "rcmax", "BMA" methods, for combining semantic similarity scores of multiple GO terms associated with protein or multiple proteins assiciated with protein cluster.

## 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

mgeneSim 11

#### **Examples**

```
## cluster1 <- c("835", "5261","241")
## cluster2 <- c("578","582")
## cluster3 <- c("307", "308", "317")
## clusters <- list(a=cluster1, b=cluster2, c=cluster3)
## mclusterSim(clusters, ont="MF", organism="human", measure="Wang")</pre>
```

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**

genes	A list of entrez gene IDs.
ont	One of "MF", "BP", and "CC" subontologies.
organism	One of "anopheles", "arabidopsis", "bovine", "canine", "chicken", "chimp", "coelicolor", "ecolik12", "ecsakai", "fly", "gondii", "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.
combine	One of "max", "average", "rcmax", "BMA" methods, for combining semantic similarity scores of multiple GO terms associated with protein or multiple proteins assiciated with protein cluster.

# verbose show progress bar or not.

#### 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

mgoSim

#### 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")
```

## Arguments

GO1	A set of go terms.
G02	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 proteins 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

termSim 13

#### See Also

goSim geneSim mgeneSim clusterSim mclusterSim

#### **Examples**

```
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")</pre>
```

termSim

termSim

#### **Description**

measuring similarities between two term vectors.

## Usage

```
termSim(t1, t2, method = c("Wang", "Resnik", "Rel", "Jiang", "Lin"),
  organism = "human", ont = "BP")
```

## Arguments

```
t1 term vector

t2 term vector

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

organism about 20 species supported, please refer to the vignettes
ont ontology
```

#### **Details**

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

#### Value

score matrix

#### Author(s)

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

14 wangMethod

# Description

Method Wang for semantic similarity measuring

# Usage

```
wangMethod(ID1, ID2, ont = "BP")
```

# Arguments

ID1	Ontology Term
ID2	Ontology Term
ont	Ontology

## Value

semantic similarity score

# Author(s)

Guangchuang Yu http://ygc.name

# **Index**

*Topic datasets IC, 8  *Topic manip clusterSim, 3 geneSim, 5 goSim, 7 mclusterSim, 10 mgeneSim, 11 mgoSim, 12  *Topic package GOSemSim-package, 2	Info_Contents_chimp_MF (IC), 8 Info_Contents_coelicolor_BP (IC), 8 Info_Contents_coelicolor_CC (IC), 8 Info_Contents_coelicolor_MF (IC), 8 Info_Contents_ecolik12_BP (IC), 8 Info_Contents_ecolik12_CC (IC), 8 Info_Contents_ecolik12_MF (IC), 8 Info_Contents_ecsakai_BP (IC), 8 Info_Contents_ecsakai_CC (IC), 8 Info_Contents_ecsakai_MF (IC), 8 Info_Contents_ecsakai_MF (IC), 8 Info_Contents_elsakai_MF (IC), 8 Info_Contents_fly_BP (IC), 8
clusterSim, 3, 3, 5, 7, 10, 12, 13 combineScores, 4	<pre>Info_Contents_fly_CC (IC), 8 Info_Contents_fly_MF (IC), 8 Info_Contents_gondii_BP (IC), 8</pre>
geneSim, 3, 4, 5, 7, 10, 12, 13 getDb, 6 getSupported_Org, 6 GOSemSim (GOSemSim-package), 2 GOSemSim-package, 2 goSim, 3-5, 7, 10, 12, 13  IC, 8 Info_Contents_anopheles_BP (IC), 8 Info_Contents_anopheles_CC (IC), 8 Info_Contents_anopheles_MF (IC), 8 Info_Contents_arabidopsis_BP (IC), 8 Info_Contents_arabidopsis_CC (IC), 8 Info_Contents_arabidopsis_MF (IC), 8 Info_Contents_arabidopsis_MF (IC), 8 Info_Contents_bovine_BP (IC), 8 Info_Contents_bovine_MF (IC), 8 Info_Contents_bovine_MF (IC), 8 Info_Contents_canine_BP (IC), 8 Info_Contents_canine_BP (IC), 8 Info_Contents_canine_CC (IC), 8	Info_Contents_gondii_BP (IC), 8 Info_Contents_gondii_CC (IC), 8 Info_Contents_gondii_MF (IC), 8 Info_Contents_human_BP (IC), 8 Info_Contents_human_CC (IC), 8 Info_Contents_human_MF (IC), 8 Info_Contents_malaria_BP (IC), 8 Info_Contents_malaria_CC (IC), 8 Info_Contents_malaria_MF (IC), 8 Info_Contents_mouse_BP (IC), 8 Info_Contents_mouse_BP (IC), 8 Info_Contents_mouse_MF (IC), 8 Info_Contents_pig_BP (IC), 8 Info_Contents_pig_BP (IC), 8 Info_Contents_pig_MF (IC), 8 Info_Contents_rat_BP (IC), 8 Info_Contents_rat_BP (IC), 8 Info_Contents_rat_MF (IC), 8 Info_Contents_rat_MF (IC), 8 Info_Contents_rhesus_BP (IC), 8 Info_Contents_rhesus_BP (IC), 8 Info_Contents_rhesus_CC (IC), 8 Info_Contents_rhesus_MF (IC), 8 Info_Contents_rhesus_MF (IC), 8
<pre>Info_Contents_canine_MF (IC), 8 Info_Contents_chicken_BP (IC), 8 Info_Contents_chicken_CC (IC), 8 Info_Contents_chicken_MF (IC), 8 Info_Contents_chimp_BP (IC), 8 Info_Contents_chimp_CC (IC), 8</pre>	Info_Contents_worm_BP (IC), 8 Info_Contents_worm_CC (IC), 8 Info_Contents_worm_MF (IC), 8 Info_Contents_xenopus_BP (IC), 8 Info_Contents_xenopus_CC (IC), 8

INDEX

```
Info_Contents_xenopus_MF (IC), 8
Info_Contents_yeast_BP (IC), 8
Info_Contents_yeast_CC (IC), 8
Info_Contents_yeast_MF (IC), 8
Info_Contents_zebrafish_BP (IC), 8
Info_Contents_zebrafish_CC (IC), 8
Info_Contents_zebrafish_MF (IC), 8
info_Contents_zebrafish_MF (IC), 8
infoContentMethod, 8

loadGOMap, 9
loadICdata, 9

mclusterSim, 3-5, 7, 10, 12, 13
mgeneSim, 3-5, 7, 10, 11, 13
mgoSim, 3-5, 7, 10, 12, 12
termSim, 13
wangMethod, 14
```