

GOSemSim

April 20, 2011

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", dr
```

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", "coelicolor", "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", "rcmax.avg" methods, for combining semantic similarity scores of multiple GO terms associated with protein or multiple proteins associated with protein cluster.

Value

sim	Semantic Similarity.
-----	----------------------

References

Wang et al.(2007) A new method to measure the semantic similarity of go terms *Bioinformatics* (Oxford, England), 23:0 1274–81, May 2007. ISSN 1460-2059 <http://www.ncbi.nlm.nih.gov/pubmed/17344234> PMID: 17344234

P W Lord et al.(2003) Semantic similarity measures as tools for exploring the gene ontology *PacificSymposiumonBiocomputing* Pacific Symposium on Biocomputing, 2003:601-12, ISSN 1793-5091 <http://www.ncbi.nlm.nih.gov/pubmed/12603061> PMID: 12603061

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")
```

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",
```

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", "ecoli12", "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.
combine	One of "max", "average", "rcmax", "rcmax.avg" methods, for combining semantic similarity scores of multiple GO terms associated with protein or multiple proteins associated with protein cluster.

Value

geneSim	Semantic similarity.
GO1	Corresponding GO terms for gene1.
GO2	Corresponding GO terms for gene2.

References

Wang et al.(2007) A new method to measure the semantic similarity of go terms *Bioinformatics* (Oxford, England), 23:0 1274–81, May 2007. ISSN 1460-2059 <http://www.ncbi.nlm.nih.gov/pubmed/17344234> PMID: 17344234

P W Lord et al.(2003) Semantic similarity measures as tools for exploring the gene ontology *PacificSymposiumonBiocomputing* Pacific Symposium on Biocomputing, 2003:601-12, ISSN 1793-5091 <http://www.ncbi.nlm.nih.gov/pubmed/12603061> PMID: 12603061

See Also

[goSim](#) [mgoSim](#) [geneSim](#) [mgeneSim](#) [clusterSim](#) [mclusterSim](#)

Examples

```
geneSim("241", "251", ont="MF", organism="human", measure="Wang")
geneSim("snR18", "YPR062W", ont="MF", organism="yeast", measure="Wang")
```

GOSemSim-package *Gene Ontology-based Sematic Similarity Measures*

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:	1.8.0
Date:	5-10-2010
biocViews:	GO, Clustering
Depends:	GO.db
Suggests:	Clustering, org.Hs.eg.db
License:	GPL Version 2

Author(s)

Guangchuang Yu

Maintainer: Guangchuang Yu <guangchuangyu@gmail.com>

See Also

[goSim](#) [mgoSim](#) [geneSim](#) [mgeneSim](#) [clusterSim](#) [mclusterSim](#)

goSim *Semantic Similarity Between Two GO Terms*

Description

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

Usage

```
goSim(GO1D1, GO1D2, ont = "MF", organism="human", measure="Wang")
```

Arguments

GO1D1	GO ID 1.
GO1D2	GO ID 2.
ont	One of "MF", "BP", and "CC" subontologies.
organism	One of "anopheles", "arabidopsis", "bovine", "canine", "chicken", "chimp", "coel-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

goSim Semantic similarity.

References

Wang et al.(2007) A new method to measure the semantic similarity of go terms *Bioinformatics* (Oxford, England), 23:0 1274–81, May 2007. ISSN 1460-2059 <http://www.ncbi.nlm.nih.gov/pubmed/17344234> PMID: 17344234

P W Lord et al.(2003) Semantic similarity measures as tools for exploring the gene ontology *PacificSymposiumonBiocomputing* Pacific Symposium on Biocomputing, 2003:601-12, ISSN 1793-5091 <http://www.ncbi.nlm.nih.gov/pubmed/12603061> PMID: 12603061

See Also

[mgoSim](#) [geneSim](#) [mgeneSim](#) [clusterSim](#) [mclusterSim](#)

Examples

```
goSim("GO:0043121", "GO:0019838", measure="Wang")
goSim("GO:0043121", "GO:0019838", ont="MF", organism="human", measure="Wang")
```

Description

These datasets are the information contents of GOterms.

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"
```

Arguments

clusters	A list of gene clusters.
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", "coelicolor", "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", "rcmax.avg" methods, for combining semantic similarity scores of multiple GO terms associated with protein or multiple proteins associated with protein cluster.

Value

simmat A Matrix of Pairwise Semantic Similarities

References

Wang et al.(2007) A new method to measure the semantic similarity of go terms *Bioinformatics* (Oxford, England), 23:0 1274–81, May 2007. ISSN 1460-2059 <http://www.ncbi.nlm.nih.gov/pubmed/17344234> PMID: 17344234

P W Lord et al.(2003) Semantic similarity measures as tools for exploring the gene ontology *PacificSymposiumonBiocomputing* Pacific Symposium on Biocomputing, 2003:601-12, ISSN 1793-5091 <http://www.ncbi.nlm.nih.gov/pubmed/12603061> PMID: 12603061

See Also

[goSim](#) [mgoSim](#) [geneSim](#) [mgeneSim](#) [clusterSim](#)

Examples

```
cluster1 <- c("835", "5261", "241", "994", "514", "517", "533")
cluster2 <- c("578", "582", "583", "400", "409", "411")
cluster3 <- c("307", "308", "317", "321", "506", "540", "378", "388", "396")
clusters <- list(a=cluster1, b=cluster2, c=cluster3)
mclusterSim(clusters, ont="MF", organism="human", measure="Wang")
```

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", comb
```

Arguments

genes	A list of entrez 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", "coelicolor", "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", "rcmax.avg" methods, for combining semantic similarity scores of multiple GO terms associated with protein or multiple proteins associated with protein cluster.

Value

simMatrix A Matrix of Pairwise Semantic Similarities

References

Wang et al.(2007) A new method to measure the semantic similarity of go terms *Bioinformatics* (Oxford, England), 23:0 1274–81, May 2007. ISSN 1460-2059 <http://www.ncbi.nlm.nih.gov/pubmed/17344234> PMID: 17344234

P W Lord et al.(2003) Semantic similarity measures as tools for exploring the gene ontology *PacificSymposiumonBiocomputing* Pacific Symposium on Biocomputing, 2003:601-12, ISSN 1793-5091 <http://www.ncbi.nlm.nih.gov/pubmed/12603061> PMID: 12603061

See Also

[goSim](#) [mgoSim](#) [geneSim](#) [clusterSim](#) [mclusterSim](#)

Examples

```
mgeneSim(c("835", "5261", "241", "994"), 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="rcmax.avg")
```

Arguments

GO1	A set 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", "rcmax.avg" methods, for combining semantic similarity scores of multiple GO terms associated with protein or multiple proteins associated with protein cluster.

Value

mgoSim Semantic similarity.

References

Wang et al.(2007) A new method to measure the semantic similarity of go terms *Bioinformatics* (Oxford, England), 23:0 1274–81, May 2007. ISSN 1460-2059 <http://www.ncbi.nlm.nih.gov/pubmed/17344234> PMID: 17344234

P W Lord et al.(2003) Semantic similarity measures as tools for exploring the gene ontology *PacificSymposiumonBiocomputing* Pacific Symposium on Biocomputing, 2003:601-12, ISSN 1793-5091 <http://www.ncbi.nlm.nih.gov/pubmed/12603061> PMID: 12603061

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")
```

Index

*Topic **datasets**

IC, 4

*Topic **manip**

clusterSim, 1

geneSim, 2

goSim, 3

mclusterSim, 5

mgeneSim, 6

mgoSim, 7

*Topic **package**

GOSemSim-package, 3

clusterSim, 1, 3–7

geneSim, 2, 2–7

GOSemSim (GOSemSim-package), 3

GOSemSim-package, 3

goSim, 2, 3, 3, 5–7

IC, 4

Info_Contents_BP_anopheles (IC), 4

Info_Contents_BP_arabidopsis
(IC), 4

Info_Contents_BP_bovine (IC), 4

Info_Contents_BP_canine (IC), 4

Info_Contents_BP_chicken (IC), 4

Info_Contents_BP_chimp (IC), 4

Info_Contents_BP_coelicolor (IC),
4

Info_Contents_BP_ecolik12 (IC), 4

Info_Contents_BP_ecsakai (IC), 4

Info_Contents_BP_fly (IC), 4

Info_Contents_BP_human (IC), 4

Info_Contents_BP_malaria (IC), 4

Info_Contents_BP_mouse (IC), 4

Info_Contents_BP_pig (IC), 4

Info_Contents_BP_rat (IC), 4

Info_Contents_BP_rhesus (IC), 4

Info_Contents_BP_worm (IC), 4

Info_Contents_BP_xenopus (IC), 4

Info_Contents_BP_yeast (IC), 4

Info_Contents_BP_zebrafish (IC), 4

Info_Contents_CC_anopheles (IC), 4

Info_Contents_CC_arabidopsis
(IC), 4

Info_Contents_CC_bovine (IC), 4

Info_Contents_CC_canine (IC), 4

Info_Contents_CC_chicken (IC), 4

Info_Contents_CC_chimp (IC), 4

Info_Contents_CC_coelicolor (IC),
4

Info_Contents_CC_ecolik12 (IC), 4

Info_Contents_CC_ecsakai (IC), 4

Info_Contents_CC_fly (IC), 4

Info_Contents_CC_human (IC), 4

Info_Contents_CC_malaria (IC), 4

Info_Contents_CC_mouse (IC), 4

Info_Contents_CC_pig (IC), 4

Info_Contents_CC_rat (IC), 4

Info_Contents_CC_rhesus (IC), 4

Info_Contents_CC_worm (IC), 4

Info_Contents_CC_xenopus (IC), 4

Info_Contents_CC_yeast (IC), 4

Info_Contents_CC_zebrafish (IC), 4

Info_Contents_MF_anopheles (IC), 4
Info_Contents_MF_arabidopsis
(IC), 4

Info_Contents_MF_bovine (IC), 4

Info_Contents_MF_canine (IC), 4

Info_Contents_MF_chicken (IC), 4

Info_Contents_MF_chimp (IC), 4

Info_Contents_MF_coelicolor (IC),
4

Info_Contents_MF_ecolik12 (IC), 4

Info_Contents_MF_ecsakai (IC), 4

Info_Contents_MF_fly (IC), 4

Info_Contents_MF_human (IC), 4

Info_Contents_MF_malaria (IC), 4

Info_Contents_MF_mouse (IC), 4

Info_Contents_MF_pig (IC), 4

Info_Contents_MF_rat (IC), 4

Info_Contents_MF_rhesus (IC), 4

Info_Contents_MF_worm (IC), 4

Info_Contents_MF_xenopus (IC), 4

Info_Contents_MF_yeast (IC), 4

Info_Contents_MF_zebrafish (IC), 4

mclusterSim, 2-4, 5, 6, 7
mgeneSim, 2-5, 6, 7
mgoSim, 2-6, 7