Package ‘flowCL’

October 7, 2014

Type Package

Title flowCL: Semantic labelling of flow cytometric cell populations

Version 1.0.0

Date 2014-03-28

Author Justin Meskas, Radina Droumeva

Maintainer Justin Meskas <jmeskas@bccrc.ca>

Depends R (>= 3.0.2), Rgraphviz, SPARQL

Description Semantic labelling of flow cytometric cell populations.

biocViews FlowCytometry, CellBiology

Suggests RUnit, BiocGenerics

License Artistic-2.0

LazyLoad yes

R topics documented:

flowCL-package *Semantic labelling of flow cytometric cell populations.*

Description

flowCL uses SPARQL to match a phenotype to a cell type from the cell ontology. If the match is not unique, then the best alternative is returned.

Details

Package:	flowCL
Type:	Package
License:	Artistic-2.0
LazyLoad:	yes
Depends:	Rgraphviz, SPARQL

flowCL executes queries against the Cell Ontology (CL), available at http://cellontology.org. The CL file is hosted on a triplestore, i.e., a database for storage and retrieval of Resource Description Framework (RDF) triples. The SPARQL endpoint at http://cell.ctde.net:8080/openrdf-sesame/repositories/CL is used to execute the SPARQL queries retrieving the correct matches from the CL. While other SPARQL endpoints can be used, users should be aware that in our case the CL file has been reasoned upon, and resulting extra inferred axioms have been added to the triplestore, providing a more complete result set.

Author(s)

Maintainer: Justin Meskas <jmeskas@bccrc.ca>

Authors: Justin Meskas, Radina Droumeva

Examples

```
# Load a pre-loaded archive. Skipping this chuck will cause flowCL to
## slowly build a new one.
data(Parents_query_archive, Parents_Names)
dir.create ( paste(getwd(),"/flowCL_results/parents_query",sep=""),
    showWarnings=FALSE, recursive=TRUE )
for (j in 1:length(Parents_Names))
    write.table(Parents_query_archive[[j]],paste(getwd(),
    "/flowCL_results/parents_query/", Parents_Names[[j]],
    sep=""), sep=",", row.names = FALSE)

# Simple two marker example
Res <- flowCL("CCR7+CD45RA+")
tmp <- Res$CCR7+CD45RA+
plot(tmp[[1]], nodeAttrs=tmp[[2]], edgeAttrs=tmp[[3]], attrs=tmp[[4]])
```

```
Res$Table

# Exact match example
Res <- flowCL("CCR7+CD45RA+CD8+", CompInfo = TRUE, OntolNamesTD = TRUE)
tmp <- Res$CCR7+CD45RA+CD8+
plot(tmp[[1]], nodeAttrs=tmp[[2]], edgeAttrs=tmp[[3]], attrs=tmp[[4]])
Res$Table

# Cell Label Example
x <-"CCR7+CD45RA+CD8+"
Res <- flowCL(x)
Res$Cell_Label[[x]][[1]]

# As a secondary way to view the results,
## see "[current directory]/flowCL_results/".
# Figures created called tree_(phenotype).pdf give the cell hierarchy
## dependent on the markers in the phenotype.
# A list of results from Res$Table are stored in listPhenotypes.csv.
```

`flowCL` *Semantic labelling of flow cytometric cell populations.*

Description

flowCL uses quaries to match a phenotype to a cell type from the cell ontology. If the match is not unique, then the best alternative is returned.

Usage

```
flowCL ( MarkerList = "HIPC", Indices = NULL, CompInfo = FALSE,
    KeepArch = TRUE, MaxHitsPht = 5, OntolNamesTD = FALSE,
    ResetArch = FALSE, VisualSkip = FALSE )
```

Arguments

`MarkerList` A list of phenotypes to query the cell ontology with (Ex. "CD3+CD4-CD8+" as a single element of the list). There is an option to use a preloaded list. This preloaded list can be accessed by pre-running data(listPhenotype_flowCL) or just inputting "HIPC" into MarkerList. This will first query all the individual markers, then will query all the common HIPC phenotypes.

`Indices` A vector of indices that dictate which of the MarkerList phenotypes will be queried. If left blank, all phenotypes in the list will be queried.

`CompInfo` A logical value that dictates if computational information is printed while the code is running. The default is FALSE.

`ResetArch` A logical value that dictates if the archive folder, "flowCL_results", is deleted before the queries take place. This will increase the simulation time, but should be done every once in a while to account for updates from the ontology. The default is FALSE.

KeepArch A logical value that dictates if the archive folder, "flowCL_results", is deleted after the queries take place. Set to FALSE to insure there is no unwanted files being stored on the hard drive. The default is TRUE.

MaxHitsPht An integer for the maximum number of cell types that are returned per phenotype queried. The default is 5.

OntolNamesTD A logical value that controls if the phenotypes in the tree diagrams(TD) are short names (ex. CD4) or the long ontology names (ex. CD4 molecule). Short names are used if OntolNamesTD is FALSE, while long ontology names are used if TRUE. The default is FALSE.

VisualSkip A logical value that controls if the visualization step is skipped or not. TRUE is for skip, while FALSE is for no skip. The default is FALSE.

Details

flowCL executes queries against the Cell Ontology (CL), available at http://cellontology.org. The CL file is hosted on a triplestore, i.e., a database for storage and retrieval of Resource Description Framework (RDF) triples. The SPARQL endpoint at http://cell.ctde.net:8080/openrdf-sesame/repositories/CL is used to execute the SPARQL queries retrieving the correct matches from the CL. While other SPARQL endpoints can be used, users should be aware that in our case the CL file has been reasoned upon, and resulting extra inferred axioms have been added to the triplestore, providing a more complete result set.

Value

A list containing N + 2 elements. Where N is the number of phenotypes queried. Each of these N elements contains information for plotting the results. The last two elements show a table and the cell labels. The table is a list of all the related information of each phenotype queried. This table is mainly for users to see the results in R. The cell labels element lists the cell labels, in order of highest hits, in a more convient way to be extracted and used by other R packages and functions.

Author(s)

Maintainer: Justin Meskas <jmeskas@bccrc.ca>

Authors: Justin Meskas, Radina Droumeva

Examples

```
# Load a pre-loaded archive. Skipping this chuck will cause flowCL to
## slowly build a new one.
data(Parents_query_archive, Parents_Names)
dir.create ( paste(getwd(),"/flowCL_results/parents_query",sep=""),
    showWarnings=FALSE, recursive=TRUE )
for (j in 1:length(Parents_Names))
    write.table(Parents_query_archive[[j]],paste(getwd(),
    "/flowCL_results/parents_query/", Parents_Names[[j]],
    sep=""), sep=",", row.names = FALSE)

# Simple two marker example
```

```
Res <- flowCL("CCR7+CD45RA+")
tmp <- Res$CCR7+CD45RA+
plot(tmp[[1]], nodeAttrs=tmp[[2]], edgeAttrs=tmp[[3]], attrs=tmp[[4]])
Res$Table

# Exact match example
Res <- flowCL("CCR7+CD45RA+CD8+", CompInfo = TRUE, OntolNamesTD = TRUE)
tmp <- Res$CCR7+CD45RA+CD8+
plot(tmp[[1]], nodeAttrs=tmp[[2]], edgeAttrs=tmp[[3]], attrs=tmp[[4]])
Res$Table

# Cell Label Example
x <-"CCR7+CD45RA+CD8+"
Res <- flowCL(x)
Res$Cell_Label[[x]][[1]]

# As a secondary way to view the results,
## see "[current directory]/flowCL_results/".
# Figures created called tree_(phenotype).pdf give the cell hierarchy
## dependent on the markers in the phenotype.
# A list of results from Res$Table are stored in listPhenotypes.csv.
```

`Parents_Names` *A list of data used to build a local archive.*

Description

A list of data used to build a local archive.

Usage

```
data(Parents_Names)
```

Author(s)

Justin Meskas

`Parents_query_archive` *A list of data used to build a local archive.*

Description

A list of data used to build a local archive.

Usage

```
data(Parents_query_archive)
```

Author(s)

Justin Meskas

`test.flowCL.connection`

A function to check that a connection exists between the user and the ontology server.

Description

A function to check that a connection exists between the user and the ontology server.

Usage

```
test.flowCL.connection()
```

Author(s)

Justin Meskas

Examples

```
test.flowCL.connection()
```

Index