

Package ‘PubMedWordcloud’

July 2, 2014

Title PubMed Word Clouds

Description create a word cloud using the abstract of publications from PubMed

Version 0.3.1

Date 2014-06-22

Author Felix Yanhui Fan <nolanfyh@gmail.com>

Imports XML, stringr, RCurl, wordcloud, tm, GOSummaries, RColorBrewer

Maintainer Felix Yanhui Fan <nolanfyh@gmail.com>

License GPL (>= 2)

URL <http://felixfan.github.io/PubMedWordcloud/>

NeedsCompilation no

Repository CRAN

Date/Publication 2014-06-23 07:24:34

R topics documented:

cleanAbstracts	2
colSets	3
editPMIDs	3
getAbstracts	4
getPMIDs	5
getPMIDsByKeyWords	5
plotWordCloud	6
pmWordCloud	7

Index	9
--------------	----------

cleanAbstracts	<i>clean data</i>
----------------	-------------------

Description

remove Punctuations, remove Numbers, Translate characters to lower or upper case, remove stop-words, remove user specified words, Stemming words.

Usage

```
cleanAbstracts(abstracts, rmNum = TRUE, tolw = TRUE, toup = FALSE,
  rmWords = TRUE, yrWords = NULL, stemDoc = FALSE)
```

Arguments

<code>abstracts</code>	output of <code>getAbstracts</code> , or just a paragraph of text
<code>rmNum</code>	Remove the text document with any numbers in it or not
<code>tolw</code>	Translate characters in character vectors to lower case or not
<code>toup</code>	Translate characters in character vectors to upper case or not
<code>rmWords</code>	Remove a set of English stopwords (e.g., 'the') or not
<code>yrWords</code>	A character vector listing the words to be removed.
<code>stemDoc</code>	Stem words in a text document using Porter's stemming algorithm.

See Also

[getAbstracts](#)

Examples

```
# Abs=getAbstracts(c("22693232", "22564732"))
# cleanAbs=cleanAbstracts(Abs)

# text="Jobs received a number of honors and public recognition."
# cleanD=cleanAbstracts(text)
```

`colSets`*plot colors***Description**

plot colors

Usage`colSets(type)`**Arguments**`type` palette names from the lists: Accent, Dark2, Pastel1, Pastel2, Paired, Set1, Set2, Set3**Examples**

```
# colors= colSets(type="Accent")
# colors= colSets(type="Paired")
# colors= colSets(type="Set3")
```

`editPMIDs`*edit PMIDs***Description**

add two sets of PMIDs together, or exclude one set PMIDs from another set of PMIDs.

Usage`editPMIDs(x, y, method = c("add", "exclude"))`**Arguments**

<code>x</code>	output of getPMIDs, or a set of PMIDs
<code>y</code>	output of getPMIDs, or a set of PMIDs
<code>method</code>	can be 'add' (default) or 'exclude'. see details.

Details

when method is 'add', PMIDs in 'x' and 'y' will be combined. when method is 'exclude', PMIDs in 'y' will be excluded from 'x'.

See Also

[getPMIDs](#)

Examples

```
# pmid1=getPMIDs(author="Yan-Hui Fan",dFrom=2007,dTo=2013,n=10)
# rm1="22698742"
# pmids1=editPMIDs(x=pmid1,y=rm1,method="exclude")

# pmid2=getPMIDs(author="Yanhui Fan",dFrom=2007,dTo=2013,n=10)
# rm2="20576513"
# pmids2=editPMIDs(x=pmid2,y=rm2,method="exclude")

# pmids=editPMIDs(x=pmids1,y=pmids2,method="add")
```

getAbstracts

get Abstracts

Description

retrieve abstracts of the specified PMIDs from PubMed.

Usage

```
getAbstracts(pmid)
```

Arguments

pmid a set of PMIDs

See Also

[getPMIDs](#)

Examples

```
# pmids=c("22693232", "22564732", "22301463", "22015308", "21283797", "19412437")
# abstracts=getAbstracts(pmids)

# pmid="22693232"
# abstract=getAbstracts(pmid)

# pmids=getPMIDs(author="Yan-Hui Fan",dFrom=2007,dTo=2013,n=10)
# abstracts=getAbstracts(pmids)
```

getPMIDs *get PMIDs using author names*

Description

retrieve PMIDs (each PMID is 8 digits long) from PubMed for author and the specified date.

Usage

```
getPMIDs(author, dFrom, dTo, n = 50)
```

Arguments

author	author's name
dFrom	start year
dTo	end year
n	max number of retrieved articles

See Also

[getAbstracts](#)

[editPMIDs](#)

Examples

```
# getPMIDs(author="Yan-Hui Fan",dFrom=2007,dTo=2013,n=10)
```

```
# getPMIDs(author="Yanhui Fan",dFrom=2007,dTo=2013,n=10)
```

getPMIDsByKeyWords *get PMIDs using Journal names and Keywords*

Description

retrieve PMIDs (each PMID is 8 digits long) from PubMed for Specific Journal, Keywords and date.

Usage

```
getPMIDsByKeyWords(keys = NULL, journal = NULL, dFrom = NULL,
  dTo = NULL, n = 10000)
```

Arguments

keys	keywords
journal	journal name
dFrom	start year
dTo	end year
n	max number of retrieved articles

See Also

[getAbstracts](#)
[editPMIDs](#)
[getPMIDs](#)

Examples

```
# getPMIDsByKeyWords(keys="breast cancer", journal="science", dTo=2013)

# getPMIDsByKeyWords(keys="breast cancer", journal="science")

# getPMIDsByKeyWords(keys="breast cancer", dFrom=2012, dTo=2013)

# getPMIDsByKeyWords(journal="science", dFrom=2012, dTo=2013)
```

plotWordCloud

PubMed wordcloud using function 'wordcloud' of package wordcloud

Description

PubMed wordcloud using function 'wordcloud' of package wordcloud

Usage

```
plotWordCloud(abs, scale = c(10, 0.2), min.freq = 1, max.words = 100,
  random.order = FALSE, rot.per = 0.35, use.r.layout = FALSE,
  colors = brewer.pal(8, "Dark2"))
```

Arguments

abs	output of cleanAbstracts, or a data frame with one colume of 'word' and one colume of 'freq'.
scale	A vector of length 2 indicating the range of the size of the words.
min.freq	words with frequency below min.freq will not be plotted
max.words	Maximum number of words to be plotted. least frequent terms dropped
random.order	plot words in random order. If false, they will be plotted in decreasing frequency
rot.per	proportion words with 90 degree rotation
use.r.layout	if false, then c++ code is used for collision detection, otherwise R is used
colors	color words from least to most frequent

Details

This function just call 'wordcloud' from package wordcloud. See package wordcloud for more details about the parameters.

See Also

[pmWordCloud](#)

Examples

```
# text="Jobs received a number of honors and public recognition."
# cleanD=cleanAbstracts(text)
# plotWordCloud(cleanD,min.freq=1,scale=c(2,1))
```

pmWordCloud

PubMed wordcloud using function 'plotWordcloud' of package GO-summaries

Description

grid based wordcloud drawing function

Usage

```
pmWordCloud(abstracts, rot.per = 0.3, max_min = c(1, 0.1), scale = 0.2,
  min.freq = 1, max.words = Inf, random.order = FALSE,
  colors = colSets(type = "Accent"), random.colors = TRUE,
  algorithm = c("circle", "leftside", "rightside"), tryfit = TRUE,
  dimensions = unit(c(1, 1), "npc"))
```

Arguments

abstracts	output of cleanAbstracts, or a data frame with one columne of 'word' and one columne of 'freq'.
rot.per	proportion words with 90 degree rotation
max_min	relative scales to adjust the size difference between largest and smallest word, by default the largest word is written with 10 times as large font than the smallest
scale	a fraction of the available space on figure that will be covered with the bounding boxes of words
min.freq	words with frequency below min.freq will not be plotted
max.words	Maximum number of words to be plotted. least frequent terms dropped
random.order	plot words in random order. If false, they will be plotted in decreasing frequency
colors	vector of colors fro the words. This vector will be extrapolated into as many colors as needed, starting with the first color for lower frequencies and ending with last color for higher frequencies.

random.colors	if true, assigns random color for the words.
algorithm	algorithm to find positions of words possible values: "circle", "leftside" and "rightside".
tryfit	if TRUE the algorithm checks if all words fit to the figure, if not it tries gradually smaller values of scale parameter until everything fits
dimensions	a two element vector of units giving the width and height of the word cloud respectively

Details

This function just call 'plotWordcloud' from package GOSummaries. See package GOSummaries for more details about the parameters.

See Also

[plotWordCloud](#)

[colSets](#)

Examples

```
# Abs=getAbstracts(c("22693232", "22564732", "22301463", "22015308", "21283797", "19412437"))
# cleanAbs=cleanAbstracts(Abs)
# pmWordCloud(cleanAbs)
```


Index

cleanAbstracts, [2](#)
colSets, [3](#), [8](#)

editPMIDs, [3](#), [5](#), [6](#)

getAbstracts, [2](#), [4](#), [5](#), [6](#)
getPMIDs, [4](#), [5](#), [6](#)
getPMIDsByKeyWords, [5](#)

plotWordCloud, [6](#), [8](#)
pmWordCloud, [7](#), [7](#)