Package ‘PubMedWordcloud’

March 1, 2019

Title 'Pubmed' Word Clouds

Description Create a word cloud using the abstract of publications from 'Pubmed'.

Version 0.3.6

Date 2019-02-28

Author Felix Yanhui Fan <nolanfyh@gmail.com>

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

Maintainer Felix Yanhui Fan <nolanfyh@gmail.com>

License GPL (>= 2)

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

RoxygenNote 6.0.1

NeedsCompilation no

Repository CRAN

Date/Publication 2019-03-01 05:30:07 UTC

R topics documented:

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

Index 8
# cleanAbstracts

**Description**

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

**Usage**

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

**Arguments**

- `abstracts`: output of `getAbstracts`, or just a paragraph of text
- `rmNum`: Remove the text document with any numbers in it or not
- `tolw`: Translate characters in character vectors to lower case or not
- `toup`: Translate characters in character vectors to upper case or not
- `rmWords`: Remove a set of English stopwords (e.g., 'the') or not
- `yrWords`: A character vector listing the words to be removed.

**See Also**

- `getAbstracts`

**Examples**

```r
# 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

<table>
<thead>
<tr>
<th>Argument</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>type</td>
<td>palette names from the lists: Accent, Dark2, Pastel1, Pastel2, Paired, Set1, Set2, Set3.</td>
</tr>
</tbody>
</table>

Examples

```r
# 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

<table>
<thead>
<tr>
<th>Argument</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>x</td>
<td>output of getPMIDs, or a set of PMIDs</td>
</tr>
<tr>
<td>y</td>
<td>output of getPMIDs, or a set of PMIDs</td>
</tr>
<tr>
<td>method</td>
<td>can be 'add' (default) or 'exclude'. see details.</td>
</tr>
</tbody>
</table>

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'.

getAbstracts

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="Yanhu 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, https = TRUE, s = 100)

Arguments

pmid a set of PMIDs
https use https instead of http
s download how many PMIDs each time

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 = 500, https = TRUE)

Arguments
author author’s name
dFrom start year
dTo end year
n max number of retrieved articles
https use https instead of http

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, https = TRUE)
plotWordCloud

Arguments

keys  keywords
journal  journal name
dFrom  start year
dTo  end year
n  max number of retrieved articles
https  use https instead of http

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.

Usage

plotWordCloud(abs, scale = c(3, 0.3), 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 column of 'word' and one column 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
plotWordCloud

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.

Examples

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

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