Package 'KDViz'

January 13, 2019

Type Package

Title Knowledge Domain Visualization

Version 1.3.1

Date 2019-01-13

Author Andres Palacios

Maintainer Andres Palacios <anfpalacioscl@unal.edu.co>

Description Knowledge domain visualization using 'mpa' co-words method as the word clustering method and network graphs with 'D3.js' library as visualization tool.

License GPL (>= 2)

Encoding UTF-8

Depends R (>= 2.10)

Imports htmlwidgets, igraph, mpa, networkD3, rvest, stringr, tm, xml2

LazyData true

NeedsCompilation no

Repository CRAN

Date/Publication 2019-01-13 17:00:03 UTC

R topics documented:

ArticleSearch	*Search articles from web*

Description

Search articles by specifying a list of key terms and a journal database.

Usage

```
ArticleSearch(keywords, webSite = "ScienceDirect", maxSize,
  saveCSV = FALSE, exportName = "BibDataOnline")
```

Arguments

keywords	a vector containing the key terms to search.
webSite	a string specifying the website where you want to search the articles. The default website is "ScienceDirect".
maxSize	the max number of article you want to get.
saveCSV	logical. If TRUE, a CSV file will be saved with the downloaded information from the bibliographic database.
exportName	a character string naming the CSV file that is going to be saved.

Details

The amount of time spent in the execution of this function depends on the internet connection and the number of articles found and the number of articles to download.

Value

A dataframe object containing the information obtained from the search of articles.

Note

Currently, SciencDirect.com is the only website from which articles information can be downloaded.

Author(s)

Andres Palacios <anfpalacioscl@unal.edu.co>

Examples

```
myBibData <- ArticleSearch(c("knowledge domain visualization"),
  webSite = "ScienceDirect", maxSize = 7, saveCSV = FALSE)

head(myBibData, 3)
```

BibToCorpus   *Convert a bibliographic database into a text corpus*

Description

Get a text corpus from a bibliographic database with a control list and help options that allow you to run a faster process of composition of corpus.

Usage

```
BibToCorpus(bibData, bibUnits = "Keywords", controlList, stopWords = TRUE,
  wordsToRemove, replaceWords)
```

Arguments

bibData	a dataframe containing information about a bibliographic database.
bibUnits	a string, the bibliographic unit to be analyzed e.g. "Title", "Keywords", "Abstract". This string must match the column name from the "bibData" dataframe.
controlList	a vector indicating the transformations and processes that will be performed during the corpus composition process. Available options: stripWhitespace for collapsing white spaces; removeNumbers for removing numbers inside texts in corpus;
stopWords	logical. If TRUE, a list of stop words will be removed from the composed corpus.
wordsToRemove	a vector of words that are desired to be removed from the composed corpus.
replaceWords	a TXT file (two columns separated by tab). One column containing the final word to be in the corpus and a second, containing the word to replace. Example: clustering cluster_analysis clustering cluster

Details

A list of stop words is provided inside the package for English language, if necessary, please visit https://sites.google.com/site/kevinbouge/stopwords-lists for a complete list of stop words in many other language, available thanks to Kevin Bouge (kevin.bouge@gmail.com)

Value

An object inheriting from VCorpus and Corpus.

Author(s)

Andres Palacios <anfpalacioscl@unal.edu.co>

See Also

[ArticleSearch](#) can be useful for creating a bibliographic information dataframe if starting from scratch.

Examples

```
data("KDVizData")
wordsToReplace <- system.file("extdata", "KDReplaceWords.txt", package = "KDViz")
wordsToRemove <- c("analysis", "data", "text", "review", "topic", "theory", "system", "protein")

myCorpus <- BibToCorpus(bibData = KDVizData, bibUnits = "Keywords",
  controlList = c("stripWhitespace", "removeNumbers"), stopWords = TRUE,
  wordsToRemove = wordsToRemove, replaceWords = wordsToReplace)
```

DTMFromCorpus *Obtain a document-term matrix from corpus*

Description

Obtain a matrix, better known as document-term matrix (DTM), where rows correspond to documents and rows to terms.

Usage

```
DTMFromCorpus(corpus, rowNames)
```

Arguments

corpus	a corpus obtained from a bibliographic database.
rowNames	a list of row names for the resulting document-term matrix to bring traceability of the names of the articles from the initial database.

Details

A quick process for obtaining a document-term matrix from a text corpus. The chosen method for weighting this matrix is the binary method, so entries of this matrix are 1 if the i-th term belongs to the j-th document and zero otherwise.

Value

a matrix object i.e. a document-term matrix, weighted by the binary method.

Note

If rowNames argument is not provided, article indexes inside document-term matrix are going to be renumbered.

Author(s)

Andres Palacios <anfpalacioscl@unal.edu.co>

Examples

```
data("KDVizData")
data("KDCorpus")

myDTM <- DTMFromCorpus(corpus = KDCorpus, rowNames = row.names(KDVizData))
```

GroupDTM   *Grouping of words with co-word method*

Description

Make partitions of an entire document-term matrix depending on the group class of words found using the KDSummary function.

Usage

```
GroupDTM(dtm, kdSummary)
```

Arguments

dtm	a document-term matrix.
kdSummary	a list resulting from [KDSummary](#) function.

Value

a list of matrices derivative from an entire document-term matrix.

Author(s)

Andres Palacios <anfpalacioscl@unal.edu.co>

See Also

For more information, see [KDSummary](#).

Examples

```
data("KDCorpus")
data("KDDTM")
mpaWords <- matrix.corpus.mpa(KDCorpus, fmin = 10, cmin = 1)
myKDSummary <- KDSummary(matrix.mpa = mpaWords, groupSize = 10)

myWordGroupDTM <- GroupDTM(dtm = KDDTM, kdSummary = myKDSummary)
```

KDCorpus   *Corpus of articles and keywords related to the topic of Knowledge*
                              domain visualization

Description

This corpus is composed by keywords inside articles related with Knowledge Domain Visualization field.

Usage

KDCorpus

Format

A corpus containing the information of terms associated to 757 articles related to the keywords `knowledge domain visualization`, `document clustering`, `co-words` and `co-citation`.

KDDTM   *Document-term matrix of articles and keywords related to the topic of*
                              Knowledge domain visualization

Description

This document-term matrix was obtained from a corpus composed by keywords inside articles related with Knowledge Domain Visualization field.

Usage

KDDTM

Format

A document-term matrix of 2486 terms associated to 555 articles related to the keywords `knowledge domain visualization`, `document clustering`, `co-words` and `co-citation`.

KDSummary   *Knowledge domain summary*

Description

Shows the main information about the grouping of words found through mpa method.

Usage

```
KDSummary(matrix.mpa, groupSize = 10, graph = TRUE)
```

Arguments

`matrix.mpa`	a list of co-occurrences and associations matrices obtained with `matrix.corpus.mpa` function.
`groupSize`	an integer that indicates the desired number of words that should belong to each group (may vary when grouping the words).
`graph`	logical. If `TRUE`, the main graph of mpa method will be plotted using `diagram.mpa` function from `mpa` package.

Details

`groupSize` argument corresponds to the size when create the groups, but it can vary according to the level of association and the amount of co-occurrences of the words.

Value

A list containing:

`clustTable`	a dataframe with the main information of the groups found.
`wordClasses`	a dataframe with words whose frequency is greater than or equal to that indicated in `matrix.mpa` argument and an indicator of the group of words to which it belongs.

Author(s)

Andres Palacios <anfpalacioscl@unal.edu.co>

See Also

mpa function from `mpa` package provides the grouping of words.

Examples

```
data("KDCorpus")
mpaWords <- matrix.corpus.mpa(KDCorpus, fmin = 10, cmin = 1)

myKDSummary <- KDSummary(matrix.mpa = mpaWords, groupSize = 10)
```

KDViz   *Knowledge Domain Visualization*

Description

Knowledge domain visualization using mpa co-words method as the word clustering method and network graphs with D3.js library as visualization tool.

Usage

```
KDViz(groupDTM, group, graph = FALSE, export = TRUE)
```

Arguments

groupDTM   a list of matrices from grouped terms and the documents in which they appear,
                obtained through GroupDTM.

group   an integer specifying the group of words that is want to be visualized.

graph   logical. If TRUE, a preview of the network graph from the chosen group will be
                plotted in R internal viewer.

export   logical. If TRUE, an html document will be saved withe the graph network using
                D3.js library.

Details

This function returns a forceNetwork from networkD3 package only if graph argument is TRUE.
By default, it will export an html document with a network graph ready for navigate in a web
browser.

Value

If graph argument is TRUE, a forceNetwork will be returned.

Author(s)

Andres Palacios <anfpalacioscl@unal.edu.co>

Examples

```
data("KDCorpus")
mpaWords <- matrix.corpus.mpa(KDCorpus, fmin = 10, cmin = 1)
myKDSummary <- KDSummary(matrix.mpa = mpaWords, groupSize = 10)
data("KDDTM")
myWordGroupDTM <- GroupDTM(dtm = KDDTM, kdSummary = myKDSummary)

KDViz(groupDTM = myWordGroupDTM, group = 1, export = TRUE)
```

KDVizData *Data about the topic of Knowledge domain visualization*

Description

This data set is bibliographic information retrieved with [`ArticleSearch`](#) funtion from this package, it contains the main information of articles related with Knowledge Domain Visualization field.

Usage

```
KDVizData
```

Format

A dataframe containing the information of `Keywords` and `URLs` from 757 articles related to the keywords `knowledge domain visualization`, `document clustering`, `co-words` and `co-citation`.

LoadArticle *Load the information stored in a bibliographic database*

Description

Load the main informaion of an article or navigate the website contained in a bibliographic database specifying the article row name.

Usage

```
LoadArticle(articleData, articleName, browser = TRUE)
```

Arguments

articleData   a dataframe containing the information of every article in the database.

articleName   a string indicating the name of the article inside the bibliographic database, this must coincide with thw row name from the article.

browser   logical. If `TRUE`, a web browser will open to visualize the article's website.

Value

information about the selected article and, possibly, the website view from it.

Note

Knowledge domain maps and original dataframe names should match, if this is not the case, please check the document-term matrix naming step in `DTMFromCorpus` step.

Author(s)

Andres Palacios <anfpalacioscl@unal.edu.co>

Examples

```
data("KDVizData")

LoadArticle(articleData = KDVizData, articleName = "A251")
```

---

matrix.corpus.mpa	*Calulation of co-occurrences matrix and matrix associations from a corpus*

Description

Complement for calculating co-ocurrences matrix and association matrix from a text corpus, using internally the function `matriz.mpa` from `mpa` package.

Usage

```
matrix.corpus.mpa(corpus, fmin = 3, cmin = 3)
```

Arguments

corpus	a corpus.
fmin	minimum frequency of appearance of key words within the corpus.
cmin	minimum match amount between words.

Value

A list containing:

MatrixA	Associations matrix.
MatrixC	Co-occurrence matrix.
Words	Vector from the different words that appears in the corpus (dictionary).
lt	Lexical table.

Author(s)

Andres Palacios <anfpalacioscl@unal.edu.co>

See Also

`matriz.mpa` function from `mpa` is used to obtain the composition of word clusters.

Examples

```
data("KDCorpus")

mpaWords <- matrix.corpus.mpa(KDCorpus, fmin = 10, cmin = 1)
```

---

read.corpus.mpa *Reading corpus object*

Description

Reads a corpus object and transforms it to be able to use `matriz.mpa` from `mpa` package.

Usage

```
read.corpus.mpa(corpus)
```

Arguments

corpus   a corpus.

Value

An object inheriting from `VCorpus` and `Corpus`.

Author(s)

Andres Palacios <anfpalacioscl@unal.edu.co>

Examples

```
data("KDCorpus")

corpusMPA <- read.corpus.mpa(KDCorpus)
```

ReadRIS *Function to read RIS files*

Description

Read the entire info from a RIS file and convert it in a dataframe object.

Usage

```
ReadRIS(risFile, saveCSV = FALSE, exportName)
```

Arguments

risFile	a RIS extension file.
saveCSV	logical. If TRUE, a CSV file of the obtained dataframe will be saved.
exportName	a character string naming the CSV file that is going to be saved.

Value

a data frame of article and its bibliographic information.

Author(s)

Andres Palacios <anfpalacioscl@unal.edu.co>

Examples

```
risData <- ReadRIS(risFile = system.file("extdata", "ScienceDirectRIS.ris",
  package = "KDViz"), saveCSV = FALSE)
```

ReplaceByList   *Replace a list of words by another*

Description

Optimized replacement of words inside a text corpus.

Usage

```
ReplaceByList(corpus, wordsFile)
```

Arguments

corpus	a VCorpus or Corpus object.
wordsFile	a TXT file (two columns separated by tab). One column containing the final word to be in the corpus and a second, containing the word to replace. Example: clustering cluster_analysis clustering cluster

Details

This function is provided as a replacing words step in a corpus composition process, this list must have a .txt file extension made by two columns, one for words of replacement and other for words to be replaced.

Value

An object inheriting from VCorpus and Corpus.

Author(s)

Camila Gongora <mcgongoraa@unal.edu.co>, Andres Palacios <anfpalacioscl@unal.edu.co>

See Also

See also [BibToCorpus](#) where this function is used internally.

Examples

```
data("KDVizData")
wordsToReplace <- system.file("extdata", "KDReplaceWords.txt", package = "KDViz")
myCorpus <- BibToCorpus(bibData = KDVizData, bibUnits = "Keywords",
  controlList = c("stripWhitespace", "removeNumbers"), stopWords = TRUE)

myNewCorpus <- ReplaceByList(myCorpus, wordsFile = wordsToReplace)
```

TermFreqByWord	*Search a specific term inside a table of word frequencies*

Description

Search a specific term inside a table of word frequencies.

Usage

```
TermFreqByWord(termFreqTable, word)
```

Arguments

termFreqTable   a table of words and frequencies.

word   a string you want to search within the table of words and frequencies.

Value

a dataframe containing a list of words which matches with the word on the search, along with its frecuencies.

Author(s)

Andres Palacios <anfpalacioscl@unal.edu.co>

See Also

A table of words and its frequencies inside a text corpus or a document-term matrix can be obtained through [TermFrequency](#) function.

Examples

```
data("KDCorpus")
termFreqTable <- TermFrequency(KDCorpus)

TermFreqByWord(termFreqTable = termFreqTable, word = "graph")
```

TermFrequency	*Table of word frequencies inside a text corpus or a document-term matrix*

Description

Function to see the total frecuencies of terms inside a text corpus or a document-term matrix object.

Usage

```
TermFrequency(x)
```

Arguments

x   a `VCorpus`, `Corpus` or `DocumentTermMatrix`.

Value

a dataframe containing a list of words and its frecuencies inside a text corpus or a document-term matrix.

Author(s)

Andres Palacios <anfpalacioscl@unal.edu.co>

Examples

```
data("KDCorpus")

termFreqTable <- TermFrequency(KDCorpus)
head(termFreqTable, 100)
```

Index