Package 'phm'

June 24, 2020

Type Package

Title Phrase Mining

Version 0.1.2

Description Functions to extract and handle commonly occurring principal phrases
obtained from collections of texts.

License GPL-3

Encoding UTF-8

LazyData true

Imports data.table (>= 1.12.8), slam (>= 0.1-46), tm (>= 0.7-7), shiny
(>= 1.4.0), utils (>= 3.6.2)

RoxygenNote 7.1.0

NeedsCompilation no

Author Ellie Small [aut, cre] (<https://orcid.org/0000-0003-1313-115X>)

Maintainer Ellie Small <esmall1@drew.edu>

Repository CRAN

Date/Publication 2020-06-24 10:00:02 UTC

R topics documented:

1

---

as.matrix.phraseDoc   *Convert a phraseDoc Object to a Matrix*

Description

Convert a phraseDoc Object to a Matrix

Usage

```
## S3 method for class 'phraseDoc'
as.matrix(x, ...)
```

Arguments

x   A phraseDoc object.

...   Additional arguments to be passed to or from methods.

Value

A matrix with phrases as rows, texts as columns, and elements containing the number of times the phrase occurs in the text.

Examples

```
tst=c("This is a test text",
      "This is a test text 2",
      "This is another test text",
      "This is another test text 2",
      "This girl will test text that man",
      "This boy will test text that man")
pd=phraseDoc(tst)
as.matrix(pd)
```

---

freqPhrases   *Display Frequent Principal Phrases*

Description

Display the most frequent principal phrases in a phraseDoc object.

Usage

```
freqPhrases(pd, n = 10)
```

Arguments

pd	A phraseDoc object.
n	Number of principal phrases to display.

Value

A vector with the n most frequent principal phrases and their frequencies.

Examples

```
tst=c("This is a test text",
      "This is a test text 2",
      "This is another test text",
      "This is another test text 2",
      "This girl will test text that man",
      "This boy will test text that man")
pd=phraseDoc(tst)
freqPhrases(pd, 2)
```

---

getDocs   *Display Frequency Matrix for Phrases*

Description

Display a frequency matrix containing all the documents that contain any of the phrases in phrs and the number of times they occur in that document.

Usage

```
getDocs(pd, phrs)
```

Arguments

pd	A phraseDoc object.
phrs	A set of phrases.

Value

A matrix with the documents and # of occurrences for the phrases in phrs.

Examples

```
tst=c("This is a test text",
      "This is a test text 2",
      "This is another test text",
      "This is another test text 2",
      "This girl will test text that man",
      "This boy will test text that man")
pd=phraseDoc(tst)
getDocs(pd, c("test text","another test text"))
```

---

phraseDoc   *phraseDoc Creation*

Description

Create an object of class phraseDoc. This will hold all principal phrases of a collection of texts that occur a minimum number of times, plus the texts they occur in and their position within those texts.

Usage

```
phraseDoc(
  txts,
  mn = 2,
  mx = 8,
  ssw = stopStartWords(),
  sew = stopEndWords(),
  sp = stopPhrases(),
  min.freq = 2,
  pp = function(phrase, freq) {     ifelse(freq < min.freq, return(FALSE),
    return(TRUE)) },
  max.phrases = 1500,
  shiny = FALSE,
  silent = FALSE
)
```

Arguments

txts	A character vector with each element the contents of a document, or a VCorpus object.
mn	Minimum number of words in a phrase.
mx	Maximum number of words in a phrase.
ssw	A set of words no phrase should start with.
sew	A set of words no phrase should end with.
sp	A set of phrases to be excluded.
min.freq	The minimum frequency of phrases to be included.
pp	Function that determines if a phrase is a principal phrase. By default, FALSE is returned if the phrase occurs less often than the number in min.freq.
max.phrases	Maximum number of phrases to be included.
shiny	TRUE if called from a shiny program.
silent	TRUE if you do not want progress messages.

Value

object of class phraseDoc.

Examples

```
tst=c("This is a test text",
      "This is a test text 2",
      "This is another test text",
      "This is another test text 2",
      "This girl will test text that man",
      "This boy will test text that man")
pd=phraseDoc(tst)
```

---

removePhrases	*Remove Phrases from phraseDoc Object*

Description

Remove a set of phrases from a phraseDoc object.

Usage

```
removePhrases(pd, phrs)
```

Arguments

pd	A phraseDoc object.
phrs	A set of phrases.

Value

A phraseDoc object with the phrases in phrs removed.

Examples

```
tst=c("This is a test text",
      "This is a test text 2",
      "This is another test text",
      "This is another test text 2",
      "This girl will test text that man",
      "This boy will test text that man")
pd=phraseDoc(tst)
removePhrases(pd, c("test text","another test text"))
```

Index