

Package ‘RcmdrPlugin.TextMining’

August 27, 2010

Type Package

Title Rcommander plugin for “tm” package

Version 0.1-0

Author Dzemil Lusija

Maintainer Dzemil Lusija <dzemill@gmail.com>

Depends R (>= 2.11.1), tcltk, tm (>= 0.5-3), XML (>= 3.1-0), Snowball (>= 0.0-7), Rgraphviz (>= 1.26.0), proxy (>= 0.4-6), Rcmdr (>= 1.5-6)

Description WARNING: This package is currently in beta status! This package provide GUI for demonstration of text mining concepts and “tm” package. It is implemented as a plugin to the R-Commander, which is based on tcl/tk. This set of dialogs can be accessed through the menu TextMining that is added to the R-Commander menus.

License Artistic-2.0

Date 2010-01-01

R topics documented:

RcmdrPlugin.TextMining-package	2
cAsoc	3
cChunks	4
cDelete	4
cDict	5
cDiss	5
cDissDoc	6
cFreq	6
cInspect	7
cInter	7
cmData	8
cPlain	8
cPlot	9
cRN	9
cRP	10
cSFT	10
cSparse	11
cStem	11

cSummary	12
cTolower	12
cUtf8	13
cWhite	13
cWords	14
cWrite	14
dirSource	15
gSource	15
rsSource	16
TDM	16
Index	17

RcmdrPlugin.TextMining-package

R-Commander plugin package that provide GUI for (tm) package

Description

The plugin adds a TextMining menu to the R-Commander. This plugin provide GUI for demontstrating functions for use in text mining using package tm

The plugin is currently a beta version. You are welcome to report any bug or inconvenience.

Details

Package:	RcmdrPlugin.TextMining
Type:	Package
Version:	1.0
Date:	2010-08-13
License:	Artistic-2.0
LazyLoad:	yes

Menu tree for RcmdrPlugin.TextMining

TextMining	- Delete corpus ...
	- Plot ...
	- Create Corpus
	- Edit Corpus
	- Transformations
	- Create DirSource...
	- Create GmaneSource...
	- Create ReutersSource...
	- Meta Data...
	- Write Corpus...
	- Make Chunks...
	- Inspect Corpus...
	- Corpus Summary...
	- To plain text documents...
	- Convert to UTF-8...
	- Remove numbers...
	- Remove punctuation...
	- Remove words...

	- Stem documents...
	- Strip whitespaces...
	- Convert to lowercase...
- Filters	- Search full text...
	- Intersection...
- TermDocumentMatrix	- Create TermDocumentMatrix...
	- Creating dictionary...
	- Find associations...
	- Find frequent terms...
	- Remove sparse terms...
- Dissimilarity	- For TermDocumentMatrix...
	- For text documents...

Remark for Windows users

John Fox, the author of R-Commander, has suggested to install R-Commander with the SDI (Single Document Interface) mode, and he also described how to change MDI to SDI, if R has been installed in the wrong mode.

Author(s)

Dzemil Lushija

Maintainer: Dzemil Lushija <dzemill@gmail.com>

References

You can find additional information about text mining package on <<http://cran.r-project.org/web/packages/tm/vignettes/tm.pdf>> and <<http://www.jstatsoft.org/v25/i05/paper>>.

cAsoc

Dialog for function findAssocs() from package (tm)

Description

This function provide interface to function `findAssocs` from **tm** package

Usage

```
cAsoc ( )
```

Details

You can also acces this function using Rcmdr menu.

From Rcmdr menu choose *TextMining* -> *TermDocumentMatrix* -> *Find associations...*

Author(s)

Dzemil Lushija <dzemill@gmail.com>

See Also

See Also [findAssocs](#)

cChunks

Dialog for function makeChunks() from package (tm)

Description

This function provide interface to function [makeChunks](#) from **tm** package

Usage

```
cChunks ()
```

Details

This function allows you to split corpus into chunks.

You can also access this function using Rcmdr menu.

From Rcmdr menu choose *TextMining -> Edit corpus -> Make Chunks...*

Author(s)

Dzemil Lushija <dzemill@gmail.com>

See Also

See also [makeChunks](#)

cDelete

Dialog for function rm() from package (base)

Description

This function provide interface to function [rm](#) from **base** package

Usage

```
cDelete ()
```

Details

From list of corpuses you can choose one to remove from workspace.

You can also access this function using Rcmdr menu.

From Rcmdr menu choose *TextMining -> Delete Corpus...*

Author(s)

Dzemil Lushija <dzemill@gmail.com>

See Also

See also [rm](#)

`cDict`*Dialog for function Dictionary() from package (tm)*

Description

This function provide interface to function [Dictionary](#) from **tm** package

Usage

```
cDict()
```

Details

With this command you can create Dictionary from list of words or from TermDocumentMatrix.

You can also acces this function using Rcmdr menu.

From Rcmdr menu choose *TextMining -> TermDocumentMatrix -> Creating dictionary...*

Author(s)

Dzemil Lushija <dzemill@gmail.com>

See Also

See also [Dictionary](#)

`cDiss`*Dialog for function dissimilarity() from package (tm)*

Description

This function provide interface to function [dissimilarity](#) from **tm** package for use with TermDocumentMatrix

Usage

```
cDiss()
```

Details

Compute the dissimilarity between documents in a TermDocumentMatrix.

You can also acces this function using Rcmdr menu.

From Rcmdr menu choose *TextMining -> Dissimilarity -> For TermDocumentMatrix...*

Author(s)

Dzemil Lushija <dzemill@gmail.com>

See Also

See also [dissimilarity](#)

`cDissDoc`*Dialog for function `dissimilarity()` from package (tm)*

Description

This function provide interface to function `dissimilarity` from **tm** package

Usage

```
cDissDoc()
```

Details

Compute the dissimilarity between two explicit documents.

You can also acces this function using Rcmdr menu.

From Rcmdr menu choose *TextMining -> Dissimilarity -> For text documents...*

Author(s)

Dzemil Lushija <dzemill@gmail.com>

See Also

See also `dissimilarity`

`cFreq`*Dialog for function `findFreqTerms()` from package (tm)*

Description

This function provide interface to function `findFreqTerms` from **tm** package

Usage

```
cFreq()
```

Details

This command allows you to find frequent terms in a TermDocumentMatrix.

You can also acces this function using Rcmdr menu.

From Rcmdr menu choose *TextMining -> TermDocumentMatrix -> Find frequent terms...*

Author(s)

Dzemil Lushija <dzemill@gmail.com>

See Also

See also `findFreqTerms`

`cInspect`*Dialog for function inspect() from package (tm)*

Description

This function provide interface to function `inspect` from **tm** package

Usage

```
cInspect ()
```

Details

Displays detailed information for a corpus.

You can also acces this function using Rcmdr menu.

From Rcmdr menu choose *TextMining -> Edit corpus -> Inspect Corpus...*

Author(s)

Dzemil Lushija <dzemill@gmail.com>

See Also

See also `inspect`

`cInter`*Dialog for function tm_intersect() from package (tm)*

Description

This function provide interface to function `tm_intersect` from **tm** package

Usage

```
cInter ()
```

Details

Perform intersection on text documents and words.

You can also acces this function using Rcmdr menu.

From Rcmdr menu choose *TextMining -> Filters -> Intersection...*

Author(s)

Dzemil Lushija <dzemill@gmail.com>

See Also

See also `tm_intersect`

cmData

Dialog for function meta() from package (tm)

Description

This function provide interface to function [meta](#) from **tm** package

Usage

```
cmData()
```

Details

Display meta data of corpus.

You can also acces this function using Rcmdr menu.

From Rcmdr menu choose *TextMining -> Edit corpus -> Meta Data...*

Author(s)

Dzemil Lushija <dzemill@gmail.com>

See Also

See also [meta](#)

cPlain*Dialog for function as.PlainTextDocument() from package (tm)*

Description

This function provide interface to function [as.PlainTextDocument](#) from **tm** package

Usage

```
cPlain()
```

Details

Converts all documents in corpus to plane text documents using [tm_map](#).

You can also acces this function using Rcmdr menu.

From Rcmdr menu choose *TextMining -> Transformations -> To plain text documents...*

Author(s)

Dzemil Lushija <dzemill@gmail.com>

See Also

See also [as.PlainTextDocument](#)

cPlot*Dialog for function plot() from package (tm)*

Description

This function provide interface to function `plot` from **tm** package

Usage

```
cPlot()
```

Details

Visualize correlations between terms of a term-document matrix.

You can also acces this function using Rcmdr menu.

From Rcmdr menu choose *TextMining -> Plot ...*

Author(s)

Dzemil Lushija <dzemill@gmail.com>

See Also

See also `plot`

cRN*Dialog for function removeNumbers() from package (tm)*

Description

This function provide interface to function `removeNumbers` from **tm** package

Usage

```
cRN()
```

Details

Remove Numbers from a Text Documents in corpus using `tm_map`.

You can also acces this function using Rcmdr menu.

From Rcmdr menu choose *TextMining -> Transformations -> Remove numbers...*

Author(s)

Dzemil Lushija <dzemill@gmail.com>

See Also

See also `removeNumbers`

`cRP`*Dialog for function `removePunctuation()` from package (tm)*

Description

This function provide interface to function `removePunctuation` from **tm** package

Usage

```
cRP ( )
```

Details

Remove all punctuation marks from a text documents in corpus using `tm_map`.

You can also acces this function using Rcmdr menu.

From Rcmdr menu choose *TextMining -> Transformations -> Remove punctuation...*

Author(s)

Dzemil Lushija <dzemill@gmail.com>

See Also

See also `removePunctuation`

`cSFT`*Dialog for function `tm_filter()` from package (tm)*

Description

This function provide interface to function `tm_filter` from **tm** package

Usage

```
cSFT ( )
```

Details

Interface to apply filter functions to corpus.

You can also acces this function using Rcmdr menu.

From Rcmdr menu choose *TextMining -> Filters -> Search full text...*

Author(s)

Dzemil Lushija <dzemill@gmail.com>

See Also

See also `tm_filter`

`cSparse`*Dialog for function `removeSparseTerms()` from package (tm)*

Description

This function provide interface to function `removeSparseTerms` from **tm** package

Usage

```
cSparse()
```

Details

Remove sparse terms from a term-document matrix.

You can also acces this function using Rcmdr menu.

From Rcmdr menu choose *TextMining -> TermDocumentMatrix -> Remove sparse terms...*

Author(s)

Dzemil Lushija <dzemill@gmail.com>

See Also

See also `removeSparseTerms`

`cStem`*Dialog for function `stemDocument()` from package (tm)*

Description

This function provide interface to function `stemDocument` from **tm** package

Usage

```
cStem()
```

Details

Stem words in a all text documents in corpus using Porter's stemming algorithm.

You can also acces this function using Rcmdr menu.

From Rcmdr menu choose *TextMining -> Transformations -> Stem documents...*

Author(s)

Dzemil Lushija <dzemill@gmail.com>

See Also

See also `stemDocument`

`cSummary`*Dialog for function `summary()` from package (base)*

Description

This function provide interface to function `summary` from **base** package

Usage

```
cSummary()
```

Details

Display summary of Corpus.

You can also acces this function using Rcmdr menu.

From Rcmdr menu choose *TextMining -> Edit corpus -> Corpus Summary...*

Author(s)

Dzemil Lushija <dzemill@gmail.com>

See Also

See also `summary`

`cTolower`*Dialog for function `tolower()` from package (base)*

Description

This function provide interface to function `tolower` from **base** package

Usage

```
cTolower()
```

Details

Translate characters from upper to lower case in all text documents in corpus.

You can also acces this function using Rcmdr menu.

From Rcmdr menu choose *TextMining -> Transformations -> Convert to lowercase...*

Author(s)

Dzemil Lushija <dzemill@gmail.com>

See Also

See also `tolower`

cUtf8*Dialog for function enc2utf8() from package (base)*

Description

This function provide interface to function [enc2utf8](#) from **base** package

Usage

```
cUtf8()
```

Details

Convert the encoding of a text documents in corpus to UTF-8.

You can also acces this function using Rcmdr menu.

From Rcmdr menu choose *TextMining -> Transformations -> Convert to UTF-8...*

Author(s)

Dzemil Lushija <dzemill@gmail.com>

See Also

See also [enc2utf8](#)

cWhite*Dialog for function stripWhitespace() from package (tm)*

Description

This function provide interface to function [stripWhitespace](#) from **tm** package

Usage

```
cWhite()
```

Details

Strip extra whitespace from all text documents in corpus. Multiple white space characters are collapsed to a single blank.

You can also acces this function using Rcmdr menu.

From Rcmdr menu choose *TextMining -> Transformations -> Strip whitespaces...*

Author(s)

Dzemil Lushija <dzemill@gmail.com>

See Also

See also [stripWhitespace](#)

cWords

Dialog for function removeWords() from package (tm)

Description

This function provide interface to function `removeWords` from **tm** package

Usage

```
cWords()
```

Details

Remove a set of words from all text documents in corpus.

You can also acces this function using Rcmdr menu.

From Rcmdr menu choose *TextMining -> Transformations -> Remove words...*

Author(s)

Dzemil Lushija <dzemill@gmail.com>

See Also

See also `removeWords`

cWrite*Dialog for function writeCorpus() from package (tm)*

Description

This function provide interface to function `writeCorpus` from **tm** package

Usage

```
cWrite()
```

Details

Write a plain text representation of a corpus to multiple files on disk corresponding to the individual documents in the corpus.

You can also acces this function using Rcmdr menu.

From Rcmdr menu choose *TextMining -> Edit corpus -> Write Corpus...*

Author(s)

Dzemil Lushija <dzemill@gmail.com>

See Also

See also `writeCorpus`

`dirSource`*Dialog for function DirSource() from package (tm)*

Description

This function provide interface to function [DirSource](#) from **tm** package

Usage

```
dirSource()
```

Details

Create corpus from DirSource.

You can also acces this function using Rcmdr menu.

From Rcmdr menu choose *TextMining -> Create Corpus -> Create DirSource...*

Author(s)

Dzemil Lushija <dzemill@gmail.com>

See Also

See also [DirSource](#)

`gSource`*Dialog for function GmaneSource() from package (tm)*

Description

This function provide interface to function [GmaneSource](#) from **tm** package

Usage

```
gSource()
```

Details

Create corpus from GmaneSource.

You can also acces this function using Rcmdr menu.

From Rcmdr menu choose *TextMining -> Create Corpus -> Create GmaneSource...*

Author(s)

Dzemil Lushija <dzemill@gmail.com>

See Also

See also [GmaneSource](#)

`rsSource`*Dialog for function ReutersSource() from package (tm)*

Description

This function provide interface to function [ReutersSource](#) from **tm** package

Usage

```
rsSource()
```

Details

Crete corpus from Reuters Source.

You can also acces this function using Rcmdr menu.

From Rcmdr menu choose *TextMining -> Create corpus -> Create ReutersSource...*

Author(s)

Dzemil Lushija <dzemill@gmail.com>

See Also

See also [ReutersSource](#)

`TDM`*Dialog for function TermDocumentMatrix() from package (tm)*

Description

This function provide interface to function [TermDocumentMatrix](#) from **tm** package

Usage

```
TDM()
```

Details

Constructs a term-document matrix or a document-term matrix.

You can also acces this function using Rcmdr menu.

From Rcmdr menu choose *TextMining -> TermDocumentMatrix -> Create TermDocumentMatrix...*

Author(s)

Dzemil Lushija <dzemill@gmail.com>

See Also

See also [TermDocumentMatrix](#)

Index

*Topic **Text mining**

- cAsoc, [3](#)
- cChunks, [3](#)
- cDict, [4](#)
- cDiss, [5](#)
- cDissDoc, [5](#)
- cFreq, [6](#)
- cInspect, [6](#)
- cInter, [7](#)
- cmData, [7](#)
- cPlain, [8](#)
- cPlot, [8](#)
- cRN, [9](#)
- cRP, [9](#)
- cSFT, [10](#)
- cSparse, [10](#)
- cStem, [11](#)
- cSummary, [11](#)
- cTolower, [12](#)
- cUtf8, [12](#)
- cWhite, [13](#)
- cWords, [13](#)
- cWrite, [14](#)
- dirSource, [14](#)
- gSource, [15](#)
- rsSource, [15](#)
- TDM, [16](#)

*Topic **package**

- cDelete, [4](#)
- RcmdrPlugin.TextMining-package,
[1](#)

as.PlainTextDocument, [8](#)

- cAsoc, [3](#)
- cChunks, [3](#)
- cDelete, [4](#)
- cDict, [4](#)
- cDiss, [5](#)
- cDissDoc, [5](#)
- cFreq, [6](#)
- cInspect, [6](#)
- cInter, [7](#)
- cmData, [7](#)

- cPlain, [8](#)
- cPlot, [8](#)
- cRN, [9](#)
- cRP, [9](#)
- cSFT, [10](#)
- cSparse, [10](#)
- cStem, [11](#)
- cSummary, [11](#)
- cTolower, [12](#)
- cUtf8, [12](#)
- cWhite, [13](#)
- cWords, [13](#)
- cWrite, [14](#)

- Dictionary, [4](#)
- DirSource, [14](#)
- dirSource, [14](#)
- dissimilarity, [5](#)

enc2utf8, [12](#)

- findAssocs, [3](#)
- findFreqTerms, [6](#)

- GmaneSource, [15](#)
- gSource, [15](#)

inspect, [6](#)

- makeChunks, [3](#)
- meta, [7](#)

plot, [8](#)

- RcmdrPlugin.TextMining
(*RcmdrPlugin.TextMining-package*),
[1](#)
- RcmdrPlugin.TextMining-package,
[1](#)
- removeNumbers, [9](#)
- removePunctuation, [9](#)
- removeSparseTerms, [10](#)
- removeWords, [13](#)
- ReutersSource, [15](#)
- rm, [4](#)

rsSource, [15](#)

stemDocument, [11](#)

stripWhitespace, [13](#)

summary, [11](#)

TDM, [16](#)

TermDocumentMatrix, [16](#)

tm_filter, [10](#)

tm_intersect, [7](#)

tm_map, [8](#), [9](#)

tolower, [12](#)

writeCorpus, [14](#)