Package 'Rtextrankr'

August 29, 2016

Type Package

Title TextRank for Korean

Version 1.0.0

Date 2016-8-25

Author Junghoon Seo

Maintainer Junghoon Seo <s3213403@gmail.com>

Description Reorder sentences for Korean text using TextRank algorithm.

License GPL-3

URL https://github.com/mikigom/Rtextrankr

Depends R (>= 3.0.2), KoNLP (>= 0.76.9)

Imports igraph (>= 1.0.1), sets (>= 1.0-16), stringi (>= 1.1.0)

Collate 'Rtextrankr.R'

RoxygenNote 5.0.1

Suggests knitr, rmarkdown

VignetteBuilder knitr

NeedsCompilation no

Repository CRAN

Date/Publication 2016-08-27 23:06:16

R topics documented:

---

build_graph *Make a graph which refers relationship of sentences. Vertex refers one sentence, and edge refers co-occurence between two sentences. Example will be shown in* R*hrefhttps://github.com/mikigom/Rtextrankrgithub.*

Description

Make a graph which refers relationship of sentences. Vertex refers one sentence, and edge refers co-occurence between two sentences. Example will be shown in github.

Usage

```
build_graph(sentences)
```

Arguments

sentences	The list of character string.

Value

The igraph graph which refers relationship of sentences.

---

co_occurence *Calculate a Jaccard Index between two sentences. Example will be shown in* R*hrefhttps://github.com/mikigom/Rtextrankrgithub.*

Description

Calculate a Jaccard Index between two sentences. Example will be shown in github.

Usage

```
co_occurence(sentence1, sentence2)
```

Arguments

sentence1	The first character string.
sentence2	The second character string.

Value

a Jaccard Index between two sentences by generalized set of noun word.

References

Tan, Pang-Ning; Steinbach, Michael; Kumar, Vipin (2005), Introduction to Data Mining, ISBN 0-321-32136-7.

`get_sentences` *Make character string list which ends with a period, from the whole text. Example will be shown in* R*hrefhttps://github.com/mikigom/Rtextrankrgithub.*

Description

Make character string list which ends with a period, from the whole text. Example will be shown in github.

Usage

```
get_sentences(text)
```

Arguments

`text`	The long character string including delimiter([.!|?|blank line]).

Value

The list of character string which ends with a period.

`sentence2table` *Convert sentence to noun word table except stopword. Example will be shown in* R*hrefhttps://github.com/mikigom/Rtextrankrgithub.*

Description

Convert sentence to noun word table except stopword. Example will be shown in github.

Usage

```
sentence2table(sentence)
```

Arguments

`sentence`	The character string.

Value

A table which refers noun word frequency, except Korean stopword.

summarize *Return list of character string of the most important sentences by Textrank algorithm. Example will be shown in* R*hrefhttps://github.com/mikigom/Rtextrankrgithub.*

Description

Return list of character string of the most important sentences by Textrank algorithm. Example will be shown in github.

Usage

```
summarize(text, count)
```

Arguments

text	The long character string including delimiter([.!|?|blank line]).
count	The number of summarized sentences.

Value

The list of character string of the most important sentences by Textrank algorithm.

References

Mihalcea, R., & Tarau, P. (2004, July). TextRank: Bringing order into texts. Association for Computational Linguistics.

table2gset *Convert sentence to word table except stopword. Example will be shown in* R*hrefhttps://github.com/mikigom/Rtextrankrgithub.*

Description

Convert sentence to word table except stopword. Example will be shown in github.

Usage

```
table2gset(table)
```

Arguments

table	A table which refers word frequency.

Value

A generalized set which refers word frequency.

xplit *Separate text into sentence list by delimiter. Example will be shown in* R*hrefhttps://github.com/mikigom/Rtextrankrgithub.*

Description

Separate text into sentence list by delimiter. Example will be shown in github.

Usage

```
xplit(text)
```

Arguments

text	The long character string including delimiter([.!|?|blank line]).

Value

The list of character string seperated by delimiter.

Index